Leveraging Machine-Vision for Activity Recognition Utilising Indoor Localisation To Support Aging-In-Place

Colin Paul Shewell, B.Sc. (Hons.)

Faculty of Computing, Engineering, and the Built Environment, Ulster University

Thesis submitted for the degree of Doctor of Philosophy

October 19, 2023

I confirm that the word count of this thesis is less than 100,000 words.

Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to express my gratitude and appreciation to my supervisors, Prof. Chris Nugent, Dr. Mark Donnelly, and Dr. Haiying Wang, who's expertise and guidance have been vital throughout the duration of this research. Not to mention their substantial patience and dedication to help support this Thesis to its final conclusion. Two additional members of the School of Computing deserve special mention, Dr. Leo Galway whose knowledge of code optimisation greatly aided in the resolution of a particularly troublesome challenge. Secondly, I would like to thank Dr. Ian Cleland who was always willing to act as a sounding board for new ideas and challenges, as well as his support as Post-Graduate Tutor within the School. I want to take this opportunity to thank my colleagues from the University of Jaén, Prof. Luis Martinez-López, Prof. Macarena Espinilla-Estévez, and Dr. Javier Medina-Quero for all their support and contributions during our collaboration.

I would like to thank my family, in particular my Mother, Barbara Shewell, for all her help and support over the last three decades. I would also like to give a very special thanks to Alison McDowell who has helped, supported, and encouraged me throughout my undergraduate and postgraduate studies as well as in wider life. Not to mention her immeasurable patience in enduring this chapter of my life! I want to express my gratitude and thanks to the close friends in my life for their unwavering support and encouragement. In particular I would like to thank Sean Maguire and Dr. Zoe Angel for their considerable support over the years and never doubting my ability to succeed.

Finally, a special mention and thanks to my Grandfather, Victor McKinney. For always encouraging and supporting me throughout my life and for inspiring my intellectual curiosity. My Grandfather's encouragement and wisdom have helped shape my journey and I will never forget his timeless advice that "*...an education is easily carried...*". As an appreciation of everything he did, I would like to dedicate this thesis to my Grandfather.

Abstract

The world is facing an unprecedented challenge where the oldest segment of society has now become the fasting growing segment of society. This is placing a large burden on existing healthcare systems who are struggling to deal with the increase in the elderly. Thus, the concept of Ambient Assisted Living to facilitate aging-in-place has come to the forefront as a potential solution to ease the burden on healthcare systems. A novel solution to this challenge using a single, wearable egocentric camera is presented. This allows a unique first-person viewpoint of the environment to be established which, through the use of fiducial markers, allows the occupant's location and current activity to be established. A study is presented assessing the technical feasibility for accurate indoor localisation to be established through the use of fiducial markers placed on key objects throughout the environment. This resulted in an effective technique to determine the current location of an occupant within an indoor environment. The tool developed within this study was then used throughout the subsequent studies as a core component of this research.

A subsequent study then sought to determine if it was possible to determine if an occupant/object interaction was a genuine interaction or a result of a cluttered environment or via navigation of the environment. The Intelligent System for Detecting Inhabitant-object Interactions (ISDII) tool was developed to determine if an interaction was genuine through the use of distance estimation to the object of interest. This study also provided a comparison between the tool developed in the previous study *vs.* an off the shelf algorithm. This study resulted in the improved performance by reducing the number of False Positives that were detected within the video stream improving precision.

A final study was carried out to not only determine the location of the occupant but to estimate their current activity. Due to the use of a wearable camera a lot of noise was introduced into the data via motion blur which resulted in missing or incorrect marker detection. Dempster-Safer theory was implemented to deal with uncertainty that was present in the data to determine the belief that an activity was being carried out. This study demonstrated the ability to reliably detect the correct

activity with an 84% success rate when tested on unreliable data.

The incorporation of these findings into the wider body of knowledge may aid in the development of future systems with the goal of solving the challenge of aging-in-place.

Abbreviations

- ADL – Activities of Daily Living
- PCRC – Pervasive Computing Research Centre
- AOTA – American Occupation Therapy Association
- TP – True Positive
- TN – True Negative
- FN – False Negative
- FP – False Positive
- ISDII – Intelligent System for Detecting Inhabitant-Objects Interactions
- GPS – Global Positioning System
- RFID – Radio-Frequency IDentification
- BLE – Bluetooth Low Energy
- AAL – Ambient Assisted Living
- FoV – Field of View
- DS – Dempster-Shafer
- QoL – Quality of Life
- ILAD – Instrumental Activities of Daily Living
- BADL – Basic Activities of Daily Living
- HMM – Hidden Markov Model
- SVM – Support Vector Machine
- ECA – Event-Condition-Action
- AmI – Ambient Intelligence
- AAL – Ambient Assisted Living
- PIR – Passive InfraRed
- ARD – Averaged Detection Ratio
- MAS – Multi-Agent System
- RSSI – Received Signal Strength Indicator
- IEEE – Institute of Electrical and Electronics Engineers
- DR – Dead Reckoning
- IMU – Inertial Measurement Unit
- UWB – Ultra-WideBand
- SIFT – Scale-Invariant Feature Transform

- RANSAC – RANdom SAmple Consensus
- CNN – Convolutional Neural Network
- LSTM – Long Short-Term Memory
- RGB – Red Green Blue
- IR – InfaRed
- R-CNN – Region-based Convolutional Neural Network
- GMM – Gaussian Mixture Model
- ANN – Artificial Neural Network
- HMM – Hidden Markov Model
- DNN – Deep Neural Network
- SPOT – Sun Small Programmable Object Technology
- IESim – Intelligent Environment Simulator
- RTSP – Real Time Streaming Protocol
- OCR – Optical Character Recognition
- ORB – Orientated FAST and Rotated BRIEF
- FAST – Features from Accelerated Segment Test
- BRIEF – Binary Robust Independent Elementary Features
- SURF – Speeded-Up Robust Features
- KNN – K-Nearest Neighbours
- DS – Dempster-Shafer
- FoV – Field of View

Chapter 1: Introduction

This Chapter offers an introduction and overview of the motivation and challenges behind this research, which investigates wearable vision-based systems for AAL. This Chapter will follow on to outline the Thesis workflow and disseminations which resulted from this work.

1.1 An Aging Population

One of the most important achievements between the 20^{th} and 21^{st} century has been the remarkable increase in life expectancy throughout the world. This has, however, resulted in the oldest group of society (aged 85 plus) becoming the most rapidly expanding sector of the population [13]. The overall burden placed on health care systems to address health problems associated with an aging society is expected to increase as this sector of the population continues to expand [13]. One potential solution to ease this burden is postulated to be through the use of a "smart environment". A smart environment can be defined as being one that is "*able to acquire and apply knowledge about the environment and its inhabitants in order to improve their experience in that environment*" [14]. It is, in the purest sense, an example of ubiquitous and pervasive computing and represents the concept of transparent "computing everywhere" [15]. It allows the support of occupants who would normally require the assistance of carers, to be supported within their own home through the incorporation of technology-based solutions [16]. It has the potential to improve quality of life and may extend the period of time a person remains living within their own home [17].

At the centre of the smart environment paradigm is wearable technology [18] enabling data to be continuously collected from a user and their immediate environment. Wearable solutions are particularly useful to support intelligent applications within smart environments where contextual information is required to provide relevant support. Contextual information includes the "*user's physical, social, emotional or informational state*" [19]. This information allows an application's behaviour to be altered to the current situation, providing task relevant information to the occupant. Beyond detecting context, there exists a number of challenges related to managing the flow and storage of data that typically originate from heterogeneous sources. Typical wearable solution applications include monitoring of vital signs, activity, social interactions, sleep patterns, along with other health indicators [20]. These parameters offer the potential for tremendous diagnostic values that were previously only possible within controlled clinical environments [21].

1.2 Role of Technology

Technology plays an important role supporting occupants within a smart environment context, not only allowing information to be collected on the occupant's current status, health, and activity but also when it comes to providing support for the occupant. This support could come in the form of simple reminders, for example, reminding an occupant when they are required to take medication or it can be used in a more holistic approach monitoring the occupant's physical, mental, and social health and making recommendations or esclating alerts to the occupant or carers/family members. The gaining ubiquity of smart phones, along with other off-the-shelf smart devices such as smart watches, has allowed this concept to flourish with the increased normalisation that the greater adaption of smart devices has brought over recent years which can be utilised to monitor occupants without them feeling as "watched" due to these devices already being present within the home. The presence and adaption of IoT devices allows for additional contextual information to be gathered about the occupants and their daily routines which can further allow support to be tailored to the occupant with the goal of providing timely support. This Thesis will focus on the use of smart glasses to aid in the monitoring of an occupant within an indoor environment via machine-vision methods. Focusing on the occupants indoor location along with their interaction with objects to perform activity recognition, this is discussed in detail within Chapter 4 where experiments are carried out to determine the feasibility of the Glass solution alongside a comparision to traditional methods.

1.3 Ethical Considerations

As mentioned in the previous section care has to be taken with regard to ethical and privacy concerns when applying a solution which relies on constant monitoring, particulary when cameras are involved. A solution which relies on continuous monitoring can result in the occupant feeling as though they are being "watched" throughout the day. This can lead to the feeling of a loss of personal space/privacy but also that of a feeling of a sense of dependancy on the technology itself [22, 23]. Additionally, there are concerns over continuous monitoring leading to a reduction in the sense of dignity for the occupant, particularly if support is needed in areas where privacy is much more of a concern, such as a bathroom or bedroom [22, 23]. While there are techniques to mitigate this issue, such only storing event data, it is important to consider the impact this may have on the occupants willingness to use the system. There is also a concern over any potential data breach which may result in personal information being made available publically. This can be mitigated through techniques such as the anonymisation of the data alongside techniques such as edge processing to keep more of the processing/data within the occupants network to minimise any potential data leak [22].

1.4 Wearable Technology

Wearable technology is rapidly becoming part of people's daily lives with the increasing popularity and uptake of devices such as smart watches and fitness trackers [24]. Within the field of healthcare wearable devices have long been used to monitor a condition and intervene if necessary [25]. The immediate detection and collection of data allows a much more real-time and accurate collection of data [26]. This allows wearable technologies to have a unique place for monitoring older people within their home, allowing for accurate, real-time measurement of their personal health and their activities [27]. This can be leveraged to aid in the support of older people living independently at home [28], allowing the possibility of remote monitoring through egocentric cameras on devices such as Google Glass [29], as presented in Figure 1.1.

1.4.1 Smart Glasses

Smart glasses are considered to be the next breakthrough in wearables [30]. Due to the popularity of smart devices a range of smart glasses are available such as, the Vuzix Blade [31], Ray-Ban Stories [32], Bose Frames [33], Snap Spectacles 3 [34], and Amazon Echo Frames [35]. This research will investigate utilising Google Glass as a sensor modality. Google Glass was initially released as the "Explorer Edition" initially in 2013, with an "Enterprise Edition" later released in 2019. “Glass Enterprise Edition” is currently on it’s second iteration – “Glass Enterprise Edition 2” [36].

(a) Top view of Google Glass.

(b) Bottom view of Google Glass.

Figure 1.1: Google Glass Explorer Edition.

Google Glass takes the form of a pair of “smart glasses” allowing a more traditional smart-phone to take the form factor of a pair of glasses. The wearer controls the Glass device through natural language commands or through a touch-pad located on the side of the device. Google Glass also contains a forward facing camera and a transparent display located in front of the wearer’s right eye, as displayed on Figure 1.2.

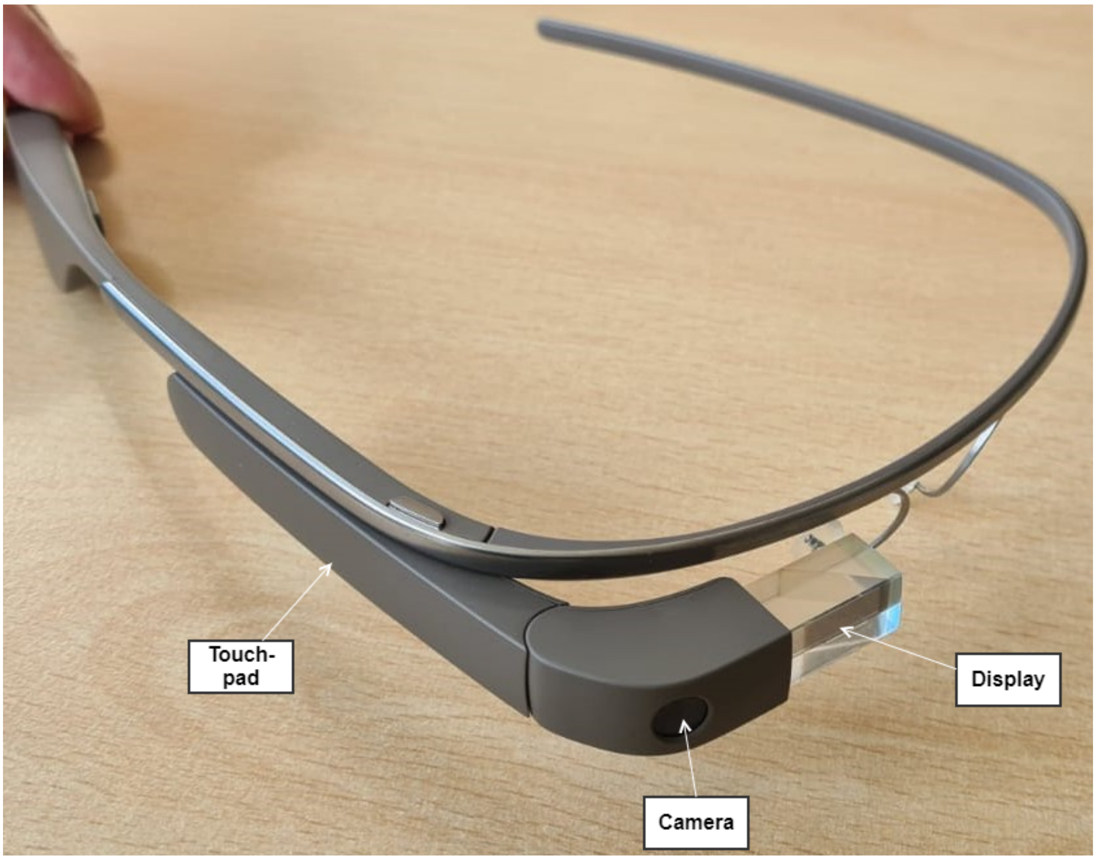


Figure 1.2: Major components of Google Glass Explorer Edition.

A number of alternative smart glass solutions have also become more common within the consumer market and it is expected to grow by 9.5% through 2028 [37]. This increase in uptake is due to additional features such as voice assistants along with improved display resolution and battery life [37] moving away from a pure AR/VR focus thus allowing businesses to see further value in the technology. There have been a number of recent developments in terms of alternative forms of smart glasses being made available. Alongside the glasses detailed previously in this section there have also been some additional offerings, such as the Xiaomi Glasses released in 2022 [38] and the EE Nreal Air AR smart glasses which were also released in 2022 [39].

1.4.2 Limitations of Wearable Technology

There are a number of limitations which must be considered when attempting to leverage wearable technology. One of the main limitations is that of battery life, many wearable devices do not have the battery capacity to run continuously for a 24 hour peroid [40] which can result in occupants being left without support as well relying on the occupant to remember to charge the device. There can also be issues with accuracy and reliability, this can be caused due to factors such as sensor placement and environmental conditions that could have an impact on the reliability of the data collected [40]. Lastly, there is also the issue of privacy and ethical concerns due to the constant monitoring that wearable devices offer. This could lead to concerns about privacy along with questions on how the data is stored and used.

1.5 Importance of Context

Traditionally technology has had a very rigid form of interaction with users, typically alerting the user with information as soon as that information becomes available or at a set time interval. The introduction of the concept of context-aware computing allows technology to detect certain contextutal information. Information such as time, date, current activity, and then adjusts its behaviour based on your context [19]. If we take the example of an older adult living alone, a context-aware system would be able to know your medication schedule and would then remind you to take your medication at an appropriate time. While also taking in to account your current activity to ensure it is not interrupting at a time where you would be unlikely to take your medication, such as when hosting a visitor. The goal of context-aware applications is to make technology more intutitive and allow more timely and relevant assistance by taking into account the current context with the goal of enhancing safety and improving support for the occupant.

Context has been defined in different ways. Brown *et al.* defined context as "*location, identities of the people around the user, the time of day, season, temperature, etc.*" [41]. Dey and Abowd defined context as the "*user's emotional state, focus of attention, location and orientation, date and time, objects, and people in the user's environment*" [19]. Although these definitions differ there are common themes that include location, time and date, current activity, and people in the user's immediate environment. Context awareness can therefore be regarded as the ability of the system to be "aware" of the user's current details, such as their location, the time of day, current social situation, and current activity. For example, assuming the user has a context aware smartphone, and they are in their bedroom and it is late at night it can be inferred that they are asleep and therefore do not wish to be disturbed by notifications. Other examples include supporting medication management where the system can remind users when their medication is due, or if they have missed a dose an automatic alert can be escalted to caregivers or family members. Additionally context aware services can offer additional advantages such as offering proactive support. If an occupant shows a decline in physical activity levels the system can prompt them to engage in some light exercise, such as walking, to aid in maintaining mobility. The inclusion of context awareness within an application offers the impression of a "smart" or "intelligent" solution, one which can seemingly anticipate the user's needs and deliver time critical information in an unobtrusive manner [42]. This type of solution is well placed for assisting the occupant in their everyday lives, in particular, to provide bespoke support to those with specific caring needs. Some types of contextual information are more important than others, in particular the location, the identity, the time, and the activity of the user [19]. Together these make up the where, who, when, and what respectively of the contextual situation.

1.5.1 Indoor Localisation

Indoor localisation is an important aspect in context aware computing [43], as determining the occupant's location is key to the system inferring the user's context. The occupant's location has been used as one of the major indicators to infer the occupant's activity as there are many areas of a building which are closely linked to the occupant's context [44]. For example, there may be core activities that take place within the kitchen that do not take place elsewhere within a living environment, such as preparing food. The occupant's location can also allow for adaptive automation within a context aware system, such as when the occupant walks into a room the lighting and heating can be adjusted to the occupant's preference. The occupant's behaviour can also be monitored by learning patterns in their daily routines and behaviour and can further aid in determining the occupant's requirements. Additionally, the tracking of an occupant's indoor location can allow for additional health monitoring by analysing their activity levels and if their routines are deviating from what is considered normal for the occupant.

1.5.2 Indoor Localisation Technologies

Research investigating the use of indoor localisation have used various technologies to determine the occupant's location [45, 46]. Some of the main approaches are the use of dense sensor placement [43], the use of active tags [47], and machine-vision techniques [48, 49]. There has been progress made within the field of location tracking technology. Common examples of these technologies include Global Positioning System (GPS), Radio-Frequency IDentification (RFID), Smart Floors, Bluetooth triangulation, and Wi-Fi fingerprinting. However, these technologies have limitations when applied to a scenario which requires accurate indoor localisation. GPS, while offering accurate localisation outdoors suffers when applied to an indoor scenario due to problems in acquiring a satellite signal when the device does not have clear access to the sky [50]. RFID offers a low cost, low power method of performing indoor localisation [50], with the RFID tags being attached to objects/persons of interest. Using RFID for the purpose of indoor localisation presents multiple challenges, such as ensuring sufficient coverage of the environment and interference from other RF emitting devices (such as Wi-Fi access points). While RFID can offer a lower cost solution to that of indoor localisation it may not be suitable for tracking an older occupant living at home. This is due to the occupant having to remember to carry an additional device that they are not accustomed to wearing on a daily basis. Smart Floors offer an accurate method of obtaining a occupant's location within an indoor environment, offering accuracy of 1cm over a 1m span [51]. While, however, a Smart Floor can offer high levels of accuracy it is an expensive method to implement, both in terms of the cost of the Smart Floor itself but also in terms of the installation costs, particularly if it will need to be retro fitting to an existing environment, which may be common within an aging-in-place situation. Smart floors are typically not suitable for an aging-in-place context due to the aforementioned cost issues being more suitable to tracking the movements of a number of people, such as in a commerical setting. One common method of determining an occupant's location is through the use of signal triangularisation, this can be achieved through the use of Bluetooth beacons [51]. Bluetooth Low-power Equipment (BLE) offers a low-cost, low-power method of providing signal triangulation within an environment [52]. The use of BLE for signal triangularisation has challenges, such as ensuring adequate coverage of an environment as well as interference, along with the maintenance of multiple beacons. Due to the need to maintain a number of receivers within the environment, coupled with the occupant being required to carry a device that they are not accustomed to carrying, can result in this method not being as suitable for monitoring an older occupant at home. Fingerprinting is an existing method to determine an occupant's location within an indoor environment [53, 54, 55]. This also relies on the environment having adequate coverage in order to reliably obtain the occupant's location as they navigate throughout the environment. Fingerprinting has some issues when being applied to an scenario of monitoring an occupant living at home. Any changes to the environment will result in having to rerun the fingerprinting process due to changes in signal strength from passing through objects in the environment.

1.5.3 Identity

The identity of the occupant is of key interest when it comes to determining the context [56, 57]. Knowing the occupant's identity allows further personalisation in order to improve the relevance of the support that can be put in place to assist the occupant within the environment. Examples of personalisation that can be used to support decisions include the occupant's medical history, their social group, as well as information on their personal daily habits. Additionally the identification of an occupant can allow personalised assistance to be offered, such as medication schedules, activity, and dietary suggestions [58]. This information can offer an insight into what is normal behavior and what is abnormal behavior for the occupant [19, 59, 57]. Once the identity of the occupant has been established it can offer a number of advantages in terms of the level of support that can be provided to the occupant. Once an occupants identity has been confirmed the system can tailor the environment to the occupants personal preferences. For example, if an occupants' typically prefers a warmer room the system can adjust the thermostat to increase the temperature to the occupants' preferred temperature. Additionally, the environment can be further tailored to the occupant's schedule, such as raising lighting to wake the occupant at their preferred time along with providing reminders for their daily schedule. The occupant's individual lifestyle can also be taken into consideration when providing support to an occupant. The occupants dietary preferences can be taken into account, providing relevant recipes or local restaurant recommendations. Additionally, the occupant's fitness goals can be taken into consideration, recommending active time depending on their activity levels throughout the day or local gyms or relevant sport clubs.

However, there can be challenges in collecting relevant data for supporting personal preferences and lifestyle factors. The occupant may feel the collection of this information to be intruding on their privacy and may not be willing to share this information or may not provide information that is fully accurate. Further challenges are differentiating between permanent changes to the occupant's routine and occasional or ad hoc changes to the occupant's routine which requires constant data collection and analysis. Lastly, collecting lifestyle/personal information can require gathering data from various sources such as wearable and environmental sensors which can introduce challenges in accuracy and consistancy across devices.

1.5.4 Time

Typically, occupants will have a certain routine that they follow which can be leveraged to determine if behaviour is abnormal [60]. For example, occupants may have a set morning routine where they carry out a typical list of activities such as a personal hygeine routine, making/eating breakfast, and getting ready for work [61]. A context aware system can use this typical routine to provide timely and relevant assistance to the occupant. Such as gently raising the lighting within the bedroom to wake the occupant, suggesting breakfast recipes, and reminding the occupant of any meetings or appointments that they may have throughout the day [62]. Additionally, a context aware system can leverage time to provide reminders at an appropriate time, such as reminders to take medication or suggesting a peroid of higher activity if their activity levels have been low throughout the day [63]. Time can also be used to directly relate to what is normal and abnormal behaviour for the occupant. For example, if they are attempting to make a meal, the time can determine if this is defined as normal behaviour (*e.g.* at 18:00) compared to abnormal behaviour (*e.g.* 03:00). This information can be used to determine if the occupant would then require further support in terms of intervention [64].The consideration of time can allow context aware systems to make timely predictions and reminders and can increase the accuracy and relevance of context aware applications and the support they can offer.

1.5.5 Activity Recognition

Activity recognition is an important factor in determining the context of a situation, there are many activities/tasks an occupant can be assumed to be doing if location, time, and identity are the only factors which are known. Therefore, it can be difficult to know what support, if any, the occupant may need at that time or indeed if the occupant is exhibiting normal or abnormal behaviour at that moment. For example, it can be classed as normal that an occupant is in the bathroom at any time of the day but with the additional information of the activity that the occupant is carrying out we can further define if that is normal or abnormal behaviour. For example, if the occupant is in the bathroom in the early hours of the morning and it was determined that they are cleaning the bathroom, this can be classed as abnormal behaviour. Furthermore, it may be defined as abnormal or normal activity for the occupant to be in the kitchen in the early hours of the morning (*e.g.*, 02:00) depending on the activity that the occupant is carrying out at the time. For example, it could be defined as normal that the occupant is in the kitchen at this time to get a glass of water yet defined as abnormal if the occupant is in the kitchen at this time and the oven is turned on along with sensors triggered for fridge/cupboards *etc.*

Activity recognition has evolved to be a critical issue in Ambient Assisted Living (AAL) as activities can give greater contextual meaning to a situation [65] and can help determine the level of independence of an occupant based on their ability to complete Activities of Daily Living (ADL) [66]. The measure of an individual to carry out their ADLs has been defined by the Katz Index of independence in ADLs [66]. This index allows the assessment of an individual's ability to carry out their ADLs in order to assess the level of assistance that may be required. Activity recognition utilises sensors placed within the environment to determine the activity via sensorised objects within the environment. Typically, this takes place through dense sensor placement (where low cost contact sensors are applied to an environment to capture data on object interaction) within an environment but more advanced techniques, such as machine-vision (using a video camera to recognise objects or occupants within the video stream) and radar (were radio waves are used to detect objects or occupants), are becoming more commonplace, as demonstrated by [67], along with combinations of multiple techniques. The constant monitoring of occupant's activities allows a more accurate evaluation of their current health status through the occupant's ability to perform ADL independently [66].

1.6 Machine-Vision

Machine-vision is a branch of computer science focused on the use of cameras and image processing to attempt to replicate human vision processing [68]. According to the Automated Imaging Association, machine-vision includes:

> *...industrial and non-industrial applications in which a combination of hardware and software provide operational guidance to devices in the execution of their functions based on the capture and processing of images.* [69]

This typically relies on digital image sensors to acquire images to allow analysis and measurements to be carried out with the end goal of informing decision making. Within the domain of AAL machine-vision can be utilised to monitor occupant's location and activities with the goal of supporting their independent living [70, 71, 72, 73]. A common method within machine-vision is that of leveraging feature points within an image [74]. These feature points will normally map to a real location within a scene allowing the comparison of the known features points of an object of interest with those feature points detected within the scene.

Traditionally utilising features within a machine-vision context consisted of three major steps, 1) feature detection, 2) feature description, and 3) feature matching [75].

Feature detection Detection is the process of detecting points of interest in an image (also known as keypoints) which can be used to uniquely identify the object of interest within the image, these features will mostly consist of edges, points, corners, and blobs [74]. Suitable features will have a well defined/localised position in the image, they should be stable under varying brightness levels and offer a high degree of repeatability in terms of detection.

Feature description Description is the process of describing the area surrounding the feature point in such a way that it will provide robustness to changes in brightness, scale, and rotation, typically resulting in a feature vector being produced for the respective feature point [75].

Feature matching Matching involves determining correlations between the known features and their descriptors of the object of interest against the features and descriptors detected in the current image.

With the advancement of machine learning techniques many "off-the-shelf" machine learning libraries now exist which can be utilised without the need to develop a unique algorithm for the application of machine-vision to AAL.

1.7 Research Challenges

There are multiple research challenges within the domain of AAL, this Thesis investigates three key issues. Firstly, the issue of a system being applied to differing environments. This traditionally requires re-training to the new environment in order to support an occupant within their own home. Secondly, the challenge of establishing the viability of determining the location of an occupant through an egocentric camera. Lastly, the challenge of establishing the current activity that the occupant is undertaking in order to provide relevant support.

Additionally, there is the issue of multiple occupancy [76, 77]. Within the context of AAL only the occupant of the environment needs support, however, False Positives (FP) can be generated which can cause irrelevant sensor events due to visitors interacting with sensors or faults in sensors. For example, FP's can be generated through care workers who may have regular check-ins with the patient or through visiting friends or family members. Further issues encountered include a lack of systems which can provide support in real-time or near real-time which is of key importance when attempting to support occupants with their ADLs. This lack of real-time support can be caused by a delay in sensor events being collected or a delay in the processing time required to make a decision.

The need for these systems to perform well in differing environments is also a challenge due to the requirement to be deployed within the occupant's own home. There can be no assumption made as to the layout of the environment or to the existence of objects that are used for location/activity recognition. Traditionally, this requires a re-training of the system to "learn" the new environment, this not only takes up time for the training process but depending on the quality of the training data that has been gathered the performance of the system will vary. There is also the issue that if any objects are moved within the environment, the system will then require training for the new room layout. The lack of a need to train for a new environment also offers a further range of benefits, such as ease of use as the system no longer has to be retrained. Which can take a considerable amount of time due to the need to perform data collection on the new environment and will allow for a faster deployment. Additionally, as the system is "pre-trained" this will typically result in a more robust and reliable model [78] which is of importance when considering the use case of supporting older adults who would require accurate and timely support.

The proposed research aims to reduce these problems though the use of Google Glass to provide a first-person (egocentric) wearable view, utilising processor off-loading and fiducial markers. Fiducial markers take the form of an object which is placed within the FoV (Field of View) of the camera to provide a unique identity to that object/scene, an example of a fiducial marker applied to an object can be seen in Figure 1.3.

Figure 1.3: A fiducial marker applied to a telephone.

A second research challenge behind the work presented was establishing the feasibility of determining the location of an occupant through a first-person wearable camera. This challenge was further compounded due to the lack of first-person perspective datasets that were available. In order to overcome this challenge a method was proposed using “key” objects within the user’s immediate environment which would then be cross-referenced against a knowledge base to determine their indoor location. One of the issues surrounding this method is the large variance that is present in common household objects, such as differing manufacturers/models of various appliances. In order to address this challenge a method was proposed using fiducial markers placed on “key” objects to allow a common and consistent method of identifying objects and thus localising the position of the occupant within the environment.

The final research challenge in this piece of work is that of determining the activity currently being carried out by the occupant. This is key for an AAL situation as the goal is to assist those in need with their ADLs in order to allow them to live in their own home independently for longer [79]. A key issue is when to determine if an object detection is due to the occupant carrying out an activity or if it is a FP due to random gaze activity or from the occupant navigating through an environment. This challenge can result in inaccurate locations being reported. In the case of determining the occupant’s activity this could result in the wrong activity being determined which could result in confusion for the occupant is support is offered for an incorrect activity.

1.8 Research Aim

This research aims to investigate the use of machine-vision to support those at home who may traditionally require assistance to carry out their ADLs through the use of improved location accuracy and activity recognition via evidential reasoning. To support this aim, a number of research quetions are posed.

1. Does the use of an egocentric wearable camera offer the ability to determine the user's indoor localisation along with additional context when detecting activities in comparison to dense sensing approaches?

2. Does the use of fiducial markers within the environment allow the easy adaption to new environments without a period of re-learning the environment?

3. Does the use of an object-distance estimation improve the rate of detection of object interaction when compared to a non-estimation approach?

4. Does the application of evidential reasoning further improve the state of the art through improving the accuracy of activity recognition?

1.9 Thesis Workflow

This Chapter offers an introduction and overview of the motivation and challenges behind this research, which investigates wearable vision-based systems for AAL. A core research challenge lies in the indoor localisation of an occupant along with the associated activity that is being carried out within the environment. Addressing these challenges using a single wearable vision-based sensor is explored, with results and challenges of such a system explored. This thesis is presented within seven Chapters. Figure 1.4. provide an illustration highlighting the relationship between these.

Chapter 2: Technology Based Approaches to Facilitate Ambient Assisted Living
This Chapter presents a range of methods and technological developments within the field of AAL, in particular, emphasis is placed on carrying out ADLs within an occupant's own home to help assist with Aging in Place.

Chapter 3: Generation of Egocentric Datasets for ADL Research
This Chapter critiques a study of work which involved generating a series of datasets to be used in a series of further studies. This Chapter includes the design of the activities and routines that will be used to generate the datasets, along

with a description of the sensors that will be used to create the data as well as a description of the environments. Details of the locations in which the dataset is gathered is presented, Ulster University Pervasive Computing Research Group and University of Jaén UJAmI SmartLab. A floor-plan layout of the respective labs is also presented along with the respective sensor locations. A detailed breakdown of the routines along with their component activities is also presented along with a technical description of the hardware used in the respective sensing technologies used within this thesis.

Chapter 4: Towards Indoor Localisation through Fiducial Marker Detection on Real-Time Video Implementing a Wearable Camera
This Chapter presents a study of work carried out to assess the technical feasibility of utilising a single wearable camera to determine occupant location via the detection of objects within the environment. A method of indoor localisation is presented through the use of an egocentric view of the environment via a single wearable camera – Google Glass. The Chapter establishes an experimental protocol to allow the feasibility of the proposed method to be assessed as a means of indoor localisation. Dense sensor placement is also used to allow a comparison of methods to be carried out to determine the success of the proposed method. A series of routines were carried out and data recorded from both the proposed system and the dense sensor placement. In order to verify if the method is applicable to multiple environments the routines were also carried out at a second test environment at the UJAmI SmartLab in the University of Jaén, Spain.

Chapter 5: Comparison of Fiducial Marker Detection and Optimising Marker and Object Detection Through Enhanced Filtering and Segmentation
This Chapter presents a study of work carried out to compare the proposed method of determining location via fiducial markers to an off the shelf method, ArUco, in order to determine the performance of the proposed system. This Chapter also presents a method to aid in determining if an occupant-object interaction is genuine or is a FP generated through the occupant navigating throughout the environment or through general gaze activity. The presented method is known as the Intelligent System for Detecting Inhabitant-object Interactions (ISDII), this is based around the observation that an occupant is generally within a known "interaction range" with the object of interest. This also takes into account the differing forms of interaction that different objects will require, a phone for instance will have a much closer interaction range than a TV. A two-stage filter was also developed for this stage in order to manage the uncertainly introduced due to missed marker

detections within the video stream.

Chapter 6: Managing Uncertainty in Activity Recognition Utilising Dempster-Shafer Theory

This Chapter presents a study of work carried out to investigate the use of DS theory. This is to minimise the uncertainty introduced through missing sensor values within a data stream when determining the activity currently being carried out by an occupant within a smart environment. The Chapter also presents the concept of DS theory and how it can be applied to an AAL context. This has been applied to the vision-based dataset within the presented work to try and minimise the effect that miss-classifications or missing sensor values have on determining the activity of interest.

Chapter 7: Conclusion

This Chapter provides a summary of the overall work presented in this Thesis along with how the overall research aims, objectives, and research questions have been addressed by this work. The Chapter also discusses the contribution to knowledge that has been made through this work as well as the limitations and directions for future work. Figure 1.4 presents the flow of work within this thesis demonstrating how each Chapter is linked.

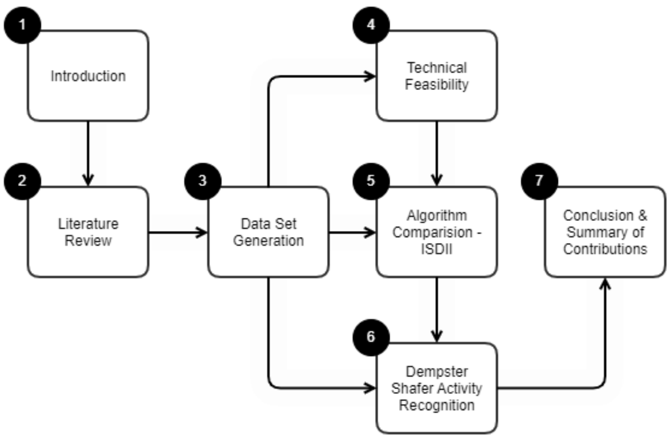


Figure 1.4: Overview of thesis showing links between Chapters that are presented. Results from Chapter 3 are used to enable the development of subsequent Chapters.

1.10 Summary of Contributions

This aim of this Thesis is to contribute to knowledge within the domain of AAL aiming to achieve the following contributions:

1. The design and implementation of a real-time vision based indoor localisation system via an egocentric camera utilising fiducial markers.

2. The design and implementation of a method to remove the need to train for each environment.

3. Benchmarking ORB and Aruco in an AAL scenario along with the development of IDSII.

4. Implementation of DS theory to that of an egocentric camera in order to correctly identify ADLs within a real world smart environment.

1.11 Research Publications

This section presents the dissemination of the work within this thesis.

Shewell, C, Nugent, C, Donnelly, M & Wang, H 2016, “Supporting Activities of Daily Living. in Active and Assisted Living: Technologies and Applications”. 1st edn, vol. 1, Institution of Engineering and Technology, Inst/Engineering & Technology, pp. 225-236. https://shop.theiet.org/active-and-assisted-living

Shewell, C, Medina-Quero, J, Espinilla, M, Nugent, C, Donnelly, M & Wang, HHY 2016, “Comparison of Fiducial Marker Detection and Object Interaction in Activities of Daily Living Utilising a Wearable Vision Sensor”, International Journal of Communication Systems.
https://doi.org/10.1002/dac.3223

Shewell, C, Nugent, C, Donnelly, M, Wang, H & Espinilla, M 2017, “Indoor Localisation Through Object Detection Within Multiple Environments Utilising a Single Wearable Camera”, Health and Technology, vol. 7, no. 1, pp. 51-60. https://doi.org/10.1007/s12553-016-0159-x

Shewell, C, Nugent, C, Donnelly, M & Wang, H 2014, “‘Wearable Technology to Aid in Activities of Daily Living” in Northern Ireland Biomedical Engineering Society Annual Spring Symposium.

Shewell, C, Nugent, C, Donnelly, M & Wang, H 2014, “Wearable computing to support activities of daily living”. in International Workshop on Ambient As-

sisted Living. vol. 8868, Lecture Notes in Computer Science , pp. 195-202. https://doi.org/10.1007/978-3-319-13105-4_30

Shewell, C, Nugent, C, Donnelly, M & Wang, H 2016, "Indoor Localisation Through Object Detection on Real-Time Video Implementing a Single Wearable Camera". in International Federation of Medical and Biological Engineering. vol. 57, Springer, pp. 1231-1236, 14th Mediterranean Conference on Medical and Biological Engineering and Computing, MEDICON 2016, 17/09/16. https://doi.org/10.1007/978-3-319-32703-7_237

1.12 Awards

The following awards were received during the course of this thesis:

- Best Poster Presentation - Faculty of Computing, Engineering, and the Built Environment Research Awards Dinner 2014.
- Santander Mobility Scholarship Award (£1,000.00) 2015 – supported a sustained, on-going collaboration with the University of Jaén, resulting in multiple publications.

Chapter 2: Technology Based Approaches to Facilitate Ambient Assisted Living

2.1 Introduction

This Chapter reviews existing research surrounding technology to promote occupants to live independently within their own homes for longer. This area of research is commonly known as AAL. AAL promotes the potential to enable inhabitants to remain within their own home for longer through the use of unobtrusive monitoring and support, allowing them to maintain an improved quality of life (QoL). Thereby reducing the burden on formal care services and delaying the potential requirement to be re-situated within full time care facilities [80].

AAL is typically realised through the use of sensor technology, which monitors the occupant's activities and to afford support with task initiation or completion, if required. AAL technologies can be used to monitor and detect anomalous behaviour, for example those relating to health related issues, such as dehydration and lack of food intake [70].

There are many methods in supporting ADL, however, they all share a common underlying methodology and with common technology, along with challenges to this area which require further research.

2.2 What is an Activity of Daily Living

Firstly, it must be defined what is an Activity of Daily Living (ADL) and what activities constitutes ADLs. Along with an overview on how these activities are supported from a traditional and technological perspective. ADL is a term used to represent the set of common tasks that comprise of one's own daily self-care requirements [81]. The ADL concept was initially proposed by Dr. Sidney Katz and his team at the Benjamin Rose Hospital in Cleveland and has since evolved into its present categorisation of activities [82]. ADL can be separated into two main categories: Basic Activities of Daily Living (BADL) and Instrumental Activities of Daily Living (IADL) [83]. To maintain BADL requires a basic competency of self-care tasks such as bathing, dressing, eating, and toileting along with the care for personal devices such as a hearing aid. IADL typically require more advanced skills as they require use of higher functions such as social skills, use of electronic devices, and the handling of money, for example. Activities catagorised as a BADL include bathing, showering, toilet/personal hygiene, eating, and sleep. IADL activities include care of others, emergency responses, child rearing, financial/health management, meal preparation, and shopping. The set of activities as defined by the American Occupation Therapy Association (AOTA) [1] for both BADL and IADL are presented in Table 2.1.

Additionally, it is important to note that the ability to carry out individual ADL may not degrade in a linear fashion, for example, activities such as bathing and dressing become increasingly impaired as conditions such as Dementia progresses whereas activities such as toileting and feeding remain relatively intact even as their condition deteriorates [84]. A possible explanation for this could be due to different cognitive areas being associated with the performance on differing ADL, rather than all ADL [84].

Table 2.1: ADL defined by the AOTA [1].

BADL	IADL
Bathing/Showering	Care of Others
Bowel and Bladder Management	Emergency Responses
Toilet Hygiene	Care of Pets
Dressing	Child Rearing
Eating	Communication Device Use
Feeding	Community Mobility
Functional Mobility	Financial Management
Personal Device Care	Health Management and Maintenance
Personal Hygiene and Grooming	Meal Preparation and Cleanup
Sexual Activity	Safety Procedures
Sleep/Rest	Shopping

As a result, the inability to carry out these activities can result in a loss of self-esteem and instill a deep sense of dependence to the person, along with a possible disturbance in family roles as partners are frequently required to assume the position of caregiver when the ability to carry out BADL/IADL is compromised [83]. Supporting people in their ADL allows them to experience a higher QoL [85], as ADL performance is directly correlated with QoL [84, 85], promotes a sense of independence and reduces the burden placed on caregivers. Additionally, a loss of independence in carrying out ADLs/IADLs has been shown to lead to a loss in autonomy and can lead to a further dependance on formal or informal care [85]. This can also result in an increase in mortality rate and their associated healthcare costs [85].

For the purpose of assessing an individual's ability to perform ADLs a number of scales have been created. One such scale is the Katz Index [86], which provides a basis for measuring, predicting, and comparing decline and recovery of the person's condition and the level of support that will be required in order for them to successfully carry out ADL [66]. Additionally, the Bristol ADL scale was developed in collaboration with caregivers to provide an assessment of people with mild dementia living in the community [87]. A person's care requirements differs depending on the degree of cognitive decline with some people losing the ability to follow instructions, or forgetting the sequence or next step of a task part way through its completion. Their ability to maintain focus on a task may decline and they often stop recognising common objects or forget how to interact with them.

However, technology offers increasing opportunities within the domain of AAL to provide increased support for those who require assistance with ADLs [88]. This is especially true in the earlier stages of cognitive decline when the person is still able to carry out tasks with a degree of independence and only require reminders or brief instructions on carrying out a task [89]. Traditionally these reminders or instructions would have to be given by caregivers, either through one to one contact or though the use of reminders left throughout the environment such as post-it notes left on items or instructions left throughout the home. Table 2.2 presents a small comparative list of such support alongside technological means of offering comparative support.

In summary, ADLs comprise essential self-care tasks and the inability for an occupant to complete these tasks can lead to a reduction in the QoL for the occupant due to a further reliance on carers or family members. Technology has been shown as a potential solution to aid those who are struggling to undertake their ADLs through offering reminders/instructions to aid in reducing the burden

Table 2.2: Comparison of differing provisions of care for ADLs from a traditional and technological perspective [2].

Traditional Care	Technological Care
Person to keep a diary for appointments	Automated calendar reminder to publish reminders/alerts to an person's smart-phone or display
Important items to be kept in the same place	Alarmed receiver attached to important objects to aid in locating
Put labels on doors/cupboards	Wearable camera to recognise and remind persons of door/cupboard contents
Place important numbers by the phone	Phone with pre-stored numbers represented by familiar faces
Place note on back of door as reminder to take keys	Door sensor to remind occupant to take keys when door is opened
Label family photographs	Facial recognition to act as a reminder
Pin a weekly timetable to the wall	Automated calendar reminder to smart-phone or display
Write reminders to lock door at night, turn off gas, put rubbish out *etc.*	Automated systems in house to take action at certain times, such as turning off cookers, locking doors, *etc.*

placed on caregivers. Addressing the issue of ADL independence is crucial to improve the lives of the occupant along with maintaining their independance and reducing the burden on the healthcare system and carers.

2.3 Leveraging Smart Environments to Support ADL

The solution of a smart environment has long been proposed as a means to ease the burden of an aging population, originally proposed by Dr. Mark Weiser in 1991 as a way to integrate computers seemlessly into our lives [15]. The development and implementation of smart environments are ongoing with contributions from industry and academia [90, 91, 92, 93, 94, 95], this research will focus on smart environments which affords occupants who would normally require the assistance of carers to be supported within their own home [96, 97, 98, 99]. This is achieved through the use of technology based solutions to allow the occupant to gain a larger degree of independence. This has been brought further into acceptance by the recent advent of consumer smart home appliances designed to be retro fitted into existing homes. A smart environment has been defined as being one that is *"able to acquire and apply knowledge about the environment and its inhabitants in order to improve their experience in that environment"* [14]. In essence, a smart environment consists of distributed technology throughout an environment and encompasses room level equipment such as lighting sensors through to object specific sensors such as automated switches. This can also take the form of sensors placed on objects within the environment, such as binary contact sensors, in order to determine the status of the occupant. Additionally, technology such as sensorised floors and/or cameras can also be installed within the environment to aid in determining the occupant's status, such as detecting falls. Chapter 3 discusses smart environments in further detail and includes visualisation of various smart environments. These technologies exist in order to gather information about an environment, which is then used to automate that environment, such as adjusting temperature via the heating system. This information can also be relayed back the person [96].

However, it should be noted that there are some limitations utilising a smart environment. The acquisition and maintenance costs of implementing a sensorised environment can be considerable. A large network of embedded sensors is normally required which results in a system that is costly to maintain, relatively obtrusive (as sensors are required on every interactable object), and sensitive to the performance of the sensors. Section 4.3.1 discusses this limitation in further detail. Additionally, there is a risk of the occupants becoming overly dependent on the technology to perform their daily tasks [100]. This over dependance on the technology can become an issue should the technology malfunction or if there is a failure in the system.

Regardless of the type of technology implemented the overall goal of a smart environment is to improve the QoL for those within the environment, in order to offer greater levels of independence, and to reduce the need for or delay institutionalisation. This is achieved through wearable and environmental sensors that allowed the facilitation of preventive care along with the monitoring of chronic conditions. It is, in the purest sense, an example of ubiquitous and pervasive computing which represents the idea of "computing everywhere", making computing and communication effectively transparent to users [15].

2.4 Technology as an Enabler

Selecting an appropriate technology to assist in supporting ADLs can be challenging with a wide range of competing technologies available. Generally, the function of these technologies can be broken down into four main applications [101]:

Ensure Safety

Fristly, ensuring the safety of the occupants is a key concern for assistive technologies. These technologies ensure that the person is not put at risk due to declining memory from conditions such as dementia, or general aging, along with the general concern from persons and family members over an occupant being left alone [102]. Some possible applications include systems such as automated door locks that will activate at a certain time [103]. It is also possible to have a camera installed that will only open the door when the person has confirmed that they know who the visitor is, facial recognition can also be used if the occupant has trouble remembering faces [104]. It is also possible to detect if the occupant may need medical assistance using technology to detect falls [105]. Where the system will detect if the occupant has fallen and contact the relevant authorities, family, or carers [106].

Improve Communication

Secondly, communication is an important factor to consider when supporting occupants. Communication technology allows the occupant to keep in contact with friends and family to help avoid them feeling alone or isolated. Simple changes can include the replacement of conventional telephones buttons with picture buttons – where a friend/relatives face is printed on the button with their number stored to assist contact [2]. Facial recognition systems can also aid in this, using door or wearable cameras the occupant can be reminded of who the visitor is along with any relevant information, such as if this is a regular carer visit or if it's a scheduled appointment.

Multi-Sensory Stimulation

Thirdly, assistive technology can aid in those with cognitive decline. These technology aids aim to relieve depression and loneliness, promoting physical well-being, along with improving relationships between people with dementia and their carers [107, 108, 109]. Examples of these include the creation of an individualised biographical reminiscence tool which provides videos, audio, and photographs from the occupant's past. With the aim to try and trigger past memories from creating a familiar sense of belonging [97].

Memory Enhancers

Lastly, assistive technology can be used to aid in memory enhancement, particularly for those with conditions which can cause increased cognitive decline [110]. Due to the high occurrence of memory related conditions amongst the elderly segment of society declining memory is a common issue [13, 110]. In order to help alleviate these issues a range of reminder technology has been developed. These include electronic calenders where appointment reminders are prompted to the occupant through a smart-phone or other display [111]. Additionally, to aid in locating important items they can have alarmed receivers attached to them to aid in locating. These can range to complex systems where the occupant's activity is recognised and assistance is then offered if the occupant is determined to be struggling to complete an activity [112].

However, it should be noted that people may react differently to differing assistive technologies. While some people may prefer a complex system, for example, one consisting of a system that monitors their medication intake and informs them accordingly, others may prefer a simple timed medicine dispenser that issues tablets at a set time each day [113]. There is also the problem that the condition of dementia, in particular, can make people apprehensive to try out new technologies with concerns over the complexity of such systems and their inability to use them, or reluctance to admit that they require assistance with ADL [102].

One potential solution to mitigate the differing responses to assistive technology is to adopt a user-centred design approach. This approach involves engaging with the target user group to better understand their requirements and their personal preferences. This can allow personalisation of the system to better fit into the users needs and lifestyle and allows the user to be involved in the decision making process. A user-centred design approach also offers longer term feedback and co-development from the user which allows the system to be further personalised and additional feature development can take place with feedback from the target users.

2.5 The Role of Ambient Intelligence in Making Environments "Smart"

One of the areas that offers a lot of opportunities within the domain of AAL is Ambient Intelligence (AmI) [114]. AmI is where the environment supports the occupants through the use of ambient sensors in place of the traditional input/output of a computer system [115, 116]. A wide range of technologies are needed to enable AmI, generally comprising of a networked range of sensors along with computational facilities to interpret the sensor information and take action based on these readings. AmI can be thought of as an amalgamation of three areas of computer science, namely — ubiquitous/pervasive computing, sensor technology, and artificial intelligence [115, 116].

The goal of activity recognition within a smart environment is to detect gradual changes in behaviour as well as atypical behaviour. Atypical behaviour could be an early sign to a change in the status of the occupant's condition or a failure in sensor equipment. The following factors need to be considered [117, 118]:

- Individuals within the environment will have differing routines and behaviour patterns. Therefore personalised classifiers for behaviour recognition are required in order to better learn what features describe an activity for an individual due [118].
- Behaviour will differ on different days of the week, such as certain activities being carried out on set days – these correlations can be learned to better determine abnormal behaviour [118].
- Occupant's routines will change over time and will display differences in the activity they undertake, the time the activity is undertaken, and the day of

the week the activity is undertaken. This will need to be taken into account when developing AmI systems [119, 118].

When activities are carried out within a similar context, *i.e.* the same time, location or carried out in a series we can then infer normal and abnormal behaviour [120]. Machine learning models can be trained to recognise what is normal behaviour through analysing labelled data [121]. Labelled data is the process of labelling each data point into predetermined categories, typically representing an activity or object interaction within the domain of AmI. The labelled data can then provide a ground truth to describe what is a normal instance of that activity. With the goal of successfully detecting normal or abnormal instances of behaviour from the occupant [122]. Additionally, time-series analysis can also be leveraged to aid in identifying correlations between days of the week and activities that the occupant carries out [123]. For example, through the occupant's historical data it may be discovered that the occupant routinly has visitors on Wednesday. Through analyising these patterns the system can make activity recommendations based on the day of the week, or conversly the system may not interrupt/disturb the occupant with low priority notifications. AmI facilitates the continuous monitoring within the home environment which can aid in early detection of deteriorating health problems or detect a worsening in chronic health conditions [124]. These may not be easily detected within a clinical environment, such as unusual behaviour such as failing to take medication correctly when there is a visitor, or failing to carrying out ADL in their daily routine.

2.5.1 Data Driven Approaches for Modelling ADL

The following section will focus on approaches taken within the domain of data driven approaches for modelling ADLs with the goal of establishing an understanding of data driven approaches. Data driven approaches rely on collecting large amounts of data to "learn" the occupant's activities and habits [115] through recognising identifiable features that make up the individual activities. Data mining is applied to the collected data in order to determine and collect patterns. Machine learning is a common technique used to reason on the data collected, this includes both supervised, unsupervised, and semi-supervised methods. Supervised methods require a set of labelled data on which to train on, for example, this consists of sensor data of the person performing activities that can then be learnt in order to recognise these patterns in real time on unknown data. Supervised and unsupervised learning form the basis of data driven approaches, where supervised approaches are given examples of activities made up from sensor data, and unsupervised methods where no example activity data is provided.

A wide range of algorithms that can be used to support supervised learning/activity recognition, these include Hidden Markov Models (HMM), Decision Trees, Support Vector Machines (SVM), Deep Learning, and Ensemble approaches which can combine a number of algorithms [65, 125, 126, 127]. An example of how an HMM could be used within activity recognition research could be a scenario to recognise walking, running, and sitting activities from an accelerometer. Firstly, data will need to be collected which is representative of the three states to detect. The features will then need to be extracted from the data to determine unique data points that can be used to identify an individual activity, some example features are mean values and standard deviation. The individual states to detect then need to be defined, in this scenrio the three states are walking, running, and sitting. This is done via training the model using a labelled dataset where data representive of each state is used so the model can "learn" the unqiue features for each state. As well as learning the transition properties which represent the transitioning from one state to another. New, unlabelled data can then be inputted into the HMM to compute what the most likely sequences of activities were undertaken that explains the observed data. K-Nearest Neighbours (KNN) can also be used for supervised learning (as well as unsupervised) which classifies activities through comparing the features of a new activity with those of its K-Nearest Neighbours within the dataset. The activity prediction is then based on the majority class amongst it's nearest neighbours [128, 129]. Naive Bayes is a probabilistic algorithm which calculates the probability that the current activity being undertaken corresponds to a known set of features for each activity [130]. It should be noted that Naive Bayes assumes that all the features used in classification are independent of each other which can be an unrealistic assumption as features in many real world datasets can be correlated [131]. Multilayer Perceptron (MLP) [132] is a neural network algorithm which consists of multiple layers of artifical neurons. An MLP is capable of learning complex patterns within a dataset which can make them suitable for tasks such as activity recognition [133]. Lastly, Random Forest is a ML algorithm that can be utilised within supervised learning, it combines multiple decision trees to make a prediction on what activity is being carried out based on the dataset. [134].

Regardless of the algorithm chosen, there is a common set of steps involved in the establishment of a representative model [65] that can accuratly capture the essential information/data for what the model is required to represent. The first step is to acquire a training dataset which will be used to train the system as to what sensor features make up an activity. This training set should be representative of the real world application that the system will be deployed within and will include labelled annotations of what the person does and when. A test dataset should also be gathered to validate the generalisation ability from the training dataset (the capacity for the model to perform well on unseen data). Once the algorithm has learnt the training dataset its performance will then be tested on the test dataset. It is common to have to repeat these steps with differing partitioning of the training and test dataset to refine the algorithms ability to detect activities while avoiding the problem of over-fitting and improve generalisation [65, 135].

Algorithms for unsupervised learning include K-Means, mixture models, Bayes networks, and KNN [65]. Unsupervised methods try to recognise and construct activities from unlabelled data, like supervised methods the first stage is to acquire a dataset but in this case unlabelled. The next stage is to aggregate and transform the sensor data into features and then model them using density estimation or clustering methods. The goal of this method is to separate the data into differing clusters, so while it may not be aware what activity represents each cluster it can determine that cluster X is a different activity to cluster Y, with the goal of identifying groups of similar data within a larger dataset. This technique is used to identify patterns within the data that may not be known within the dataset. Clustering is achieved via utilising a distance measurement within the data, such as Hamming or Euclidean distance, in order to determine how similiar each respective data point is. The data is then arranged into clusters depending on their distance measurement in order to discover hidden patterns within the dataset. Additionally, density estimation can be utilised for understanding the underlying data distribution, which involves estimating the liklihood of data points with the aim of determining how the data is distrubuted within the dataset. One major challenge using a data driven approach, whether supervised or unsupervised, can be the requirement to obtain a vast amount of data for training purposes. This issue is further compounded if attempting to incorporate video and audio data into the reasoning process [114].

Semi-supervised method are a hybrid approach which combines supervised and unsupervised learning[136]. This method uses both labelled and unlabelled data, this is normally used when the labelled data is not comprehensive enough to produce an accurate model or when accurately labelling the data would be too time intensive to be feasible. A semi-supervised method uses a limited set of labelled data to train itself as per a supervised method, resulting in a “partially” trained model. This partially trained model then labels the unlabelled data, known as pseudo-labelled data. The labeled and pseudo-labelled datasets are then combined allowing both the descriptive aspects of supervised learning to be combined with the predictive aspects of unsupervised learning.

2.5.2 Knowledge Driven Approaches for Modelling ADL

Knowledge driven based systems rely on a series of rules that determines if an activity is being carried out. Knowledge based systems are designed to utilise domain specific expertise to make decisions allowing them to make complex decisions in order to solve problems. One advantage of knowledge driven system is that it allows you to separate the activity detection from the supporting system thus allowing rules to be reused within the domain with only minimal customisation to the unique needs that the occupant may require [137, 138]. There are differing methods of knowledge representation within smart environments, however, there is currently no accepted standard within the domain.

Event-Condition-Action Systems

One form of knowledge driven systems is the Event-Condition-Action (ECA) which is an architectural pattern for representing context awareness [139, 140]. This method consists of three modules, an Event, Condition, and action modules — the Event module is responsible for gathering contextual information such as sensor data, the Condition module is responsible for the rules and the Action module is responsible for executing the associated action with each rule [141, 140].

These rules take the form of:

On (event expression)
If (condition)
Do (action)

The reading of these rules are as follows: *On* detecting a certain event check *If* a condition is true and if so then *Do* the specified action [141, 140]. The *On* is defined as the rule trigger which is determined to be true if an event occurs that matches the event expression defined in *On*. The *If* section of the rule defines a conditional statement that has to be evaluated to true in order to trigger the final statement of the rule which specifies what action needs to be taken [115]. A simple example of this would be:

On (bedSensor == TRUE)
If (TV == ON)
Do (TV = OFF)

This reads as *On* detection that someone is in bed and *If* the TV is still on, then *Do* turn the TV off. An example of a traditional implementation of an ECA system is through the use of HomeRuleML [142] which is an XML based schema to represent rules within a smart environment. HomeRuleML specifies a list of sensor IDs along with a conditional value which when true allows the system to determine if the conditions of an activity/rule have been specified along with an action statement for each rule.

Ontology Systems

Another form of knowledge driven AmI is through the use of an ontology based system, which allows taxonomies and relationships between concepts to be defined [143, 144]. Ontologies offer several advantages over traditional forms of knowledge representation, a well-defined ontology allows knowledge sharing a re-use [145], declarative semantics allow multiple policies to support context detection [146], and ontologies also provide complex inference mechanisms [147, 148].

Limitations

However, it should be noted there are a number of limitations with implementing and maintaining knowledge based systems. Firstly, acquiring and maintaining the knowledge required for these systems can be a challenge due to the requirement to have a domain-expert to provide the knowledge [149]. Gathering the required knowledge can be a time consuming and costly process, particuraly as knowledge will change over time which will require the sytem to be regularly updated with new domain knowledge [150]. There is also the challenge of how to best represent the domain knowledge in a format that is understandable to machines as well as humans, particularly when ensuring transparancy in the decision being made by the system.

2.5.3 Context Driven Approaches for Modelling ADL

When developing context-aware applications the users contextual situation is key to supporting the occupant. Context-aware approaches rely on information such as the date, time, occupant's location, roles of people present, as well as known objects [115]. As introduced in Chapter 1, Dey and Abowd define context as:

> *"...any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves..."* [19]

Contextual Categories

Context can be used to represent two main categories within AAL — user-centric context and environmental context [115].

User-Centric Context

User-centric context revolves around the occupant's background, current behaviour, and emotional state [115]. Background factors such as a user's interests or medical conditions can have an effect on their current context or the actions that will need to be taken. For example, should the person be diabetic, support will be required to ensure that blood glucose levels are checked at regular intervals and before meals. Current behaviour will factor in variables such as the person's current activity. If they are currently involved in an activity viewed as high importance, such as a discussion with their physician, then the person will not be notified/prompted with items that are considered low importance. Lastly the person's emotional state is taken into account when determining user-centric context from multimodal sensors and analysis of user features, such as voice or tremors, as you may not want to further frustrate the person if they are in a poor emotional state. Sokullu *et al.* [151] developed a system which offered reminders depending on the serverity of the abnormal behaviour and the occupant's context. An example would be if the occupant had left the bathroom tap running, the system determined that a "mild" reminder would be suffice when the occupant was near the tap as it not categorised as an immediate danger to the occupant.

Environmental Context

Environmental context revolves around the occupant's physical, social, and computational surroundings [56]. Physical factors include variables such as the current time, the occupant's physical location, and temperature. These can have a large impact on the type of support that is required, as certain activities will only take place within a certain time frame that is in a certain location — such as cooking dinner which would normally be within the hours of 17:00 – 19:00 within the kitchen [59]. Social factors will involve variables such as surrounding people, for example, if the occupant is currently being visited by carers or family members they will most likely not wish to be disturbed [113]. The final environmental factor is that of the person's immediate computational surroundings, this takes into consideration equipment such as sensors or displays that are within close proximity. In order to avoid problems such as sending notifications to displays which the person may not be able to see or that are in a different room to the person's current location [115]. Cha *et al.* [63] developed a system which investigated the contextual factors relevant to interruptibility when providing a reminder to an occupant, taking into account the environmental context. They found that when the occupant was co-located with other occupants who are all undertaking the same activity then the interruptibility of the occupant is dependant on their engagement with the activity and the urgency.

Dey and Abowd further define the categories for context aware applications [19] as the following:

Presentation

Is the ability to display information that is relevant to the user, including contextual information, and not just a list of information that requires further user interaction. Presentation involves how the relevant contextual information is presented to the occupant to ensure it is understandable, timely, and useful for the occupant. An example of this would be a mobile device that allows the display of friends or family member's location and an awareness of their activity. For a context aware navigation system could display an indoor floorplan with route information and turn by turn instructions based on the occupant's current location. It is an amalgamation of Schilit's proximate selection [152] and Pascoe's notion of presenting context [153].

Execution

Involves adapting the environment with additional information by associating particular data with a particular context through the actions and decisions taken by the system. This involved the systems' ability to adapt it's behaviour in a dynamic fashion to respond to the occupants' current situation. An example of this would be alerting a person when a visitor is within a certain distance of their home. A context-aware system the execution layer could send medication reminders or adjust the occupant's activity levels/exercise routines based on the occupant's current health data or their current level of physical activity. Execution is based on Schilit's context-triggered actions, and Pascoe's contextual adoption.

Tagging

Tagging is the process in which the system records the actions and times that they were carried out. This involves associating (or labelling) contextual information with the actions or objects that are relevant to the occupant's activity. This can then be used later to help determine the person's behavioural habits and also to help determine if chronic conditions are deteriorating. This information can also be used in a data mining approach in order to recognise patterns, such as online shopping applications which tag user preferences/purchases allowing the system to suggest other products in the future which are based on the user's preferences.

This definition of contextual information allows a system's behaviour to be personalised to the users' current situation through the use of the presentation, execution, and tagging categories. The presentation category provides a mechanism in which the users can percieve and understand the data and contextual information. The execution category provides a mechanisim in which the system can make context-aware actions, and tagging allows a mechanism to aid in the organisation of the contextual information to assist in providing enhanced services and experiences.

In practice there are certain contextual variables that are more important than others. Typically these include: location, identity, time, and the person's current activity. For example, location permits nearby objects, people, and activities to be determined. Their identity then allows other background information to be inferred, such as contact details, birth date, list of friends, and relationships to other people within their environment. A computer system that has knowledge of context is therefore able to sense, and react based on the person's requirements within an environment in order to improve their QoL.

2.6 Ethical and Security Issues

One aspect of AAL (using technologies to enable inhabitants to remain within their own home for longer through the use of unobtrusive monitoring and support) that is unavoidable is the ethical and security concerns of such intrusive technologies. This is a deep and wide ranging issue and can only be covered briefly within the scope of this chapter.

2.6.1 Ethical Issues

The implementation of the technology discussed previously in this Chapter raises several ethical issues pertaining to the privacy and dignity of the occupants. Their safety outside of supervised care, along with the quality of care that can be provided. Other complex issues arise such as identification of who is responsible for the occupant's well-being as they are no longer solely being treated by medical staff/carers. There is also the issue of who will be responsible for the maintenance of such a system, will carers be called on to both manage the person and to maintain the system [154]. The issue of front facing cameras, such as that on smart glasses, is also a pressing privacy concern. Previous attempts to maintain privacy include image blurring in sensitve areas or data anonymisation [155]. Additionally, advancements within edge computing has offered increased privacy through a shift in processing and data storage from the cloud to the end-users or near-user edge devices [156].

Privacy

One issue is that of the occupant's privacy, due to the range of sensing technology that is used within AAL, a range of data is continuously being collected. There is a challenge in balancing the level of technology and data that is collected to provide support while respecting the occupant's right to privacy. Egocentric cameras which provide a first-person view of the environment can raise particular privacy concerns. Some possible mitigation strategies could be to blur the images in sensitive areas or to only store event data and not store the raw vision data.

Unsupervised Safety

The goal of AAL systems is to allow occupants to live independently at home but there is a concern about the occupants safety when they are no longer wtihin supervised care. Care and consideration needs to be taken to ensure that occupants remain safe when not under supervised care, such as ensuring that they remain able to contact emergency services.

Responsibility

Determining responsibility for an occupant's wellbeing becomes more complex with the adoption of AAL systems. This is due to responsibility being no longer the sole domain of the medical staff caring for the occupant. The responsibilities of those who are developing, installing, and maintaining these system will need to be clarified within future work.

2.6.2 Security Issues

One major concern within the domain of AAL is that of security, as a smart environment will be storing personal information about an occupant, which may include medical data, therefore security is an utmost priority [157]. While there is legislation in place to provide guidelines [158] on the access and usage of medical data, security is still considered a major issue. An overview of some of the major threats is presented in [159].

Personal Information

While legislation does exist within certain geopolitical areas, such as GDPR in Europe, safeguarding this data is crucial. Ensuring that there are secure authentication steps are put in place to avoid unauthorised users from tampering with the system or to prevent them from gaining access to private data.

Encryption

Data that is transmitted between sensors, devices, or servers must be encrypted to mitigate against interception or data leaks. This can aid in preserving the confidentiality and integrity of the occupant's personal data.

Physical Security

While the importance of authorisation and encryption cannot be understated it is important to also consider the physical security of the system. Unathorised physical access can led to malicious actors tampering with the system or gaining access to private data.

In summary, AAL systems face a range of ethical and security challenges which comprises issues such as privacy and safety along with corresponding security challenges such as encryption. To ensure the future adoption and success of AAL systems a holistic approach will be necessary, addressing both privacy and security in terms of technological and physical.

2.7 Emerging Trends in Sensor Types

Sensors used to support ADL range from discrete state sensors to those that continuously record data. There are a range of methods which can be used to support ADLs within their own home. Table 2.3 presents a summary of commonly used sensor types, as reported in [160, 3]. One such method is Dense sensor placement

[161]. A dense sensor system consists of a large array of lost cost sensors (binary contact sensors in this implementation) which are placed on every object that the inhabitant may interact with. This allows their location to be determined based on their object interaction. Contact/pressure sensors will record binary state data to inform whether an object has been interacted with, such as a door or bed respectively. Accelerometers and gyroscopes [162] are normally combined to assess how active the occupant is, along with other measures such as gait analysis. There is also a range of sensing technology used for indoor location. Indoor location can be key to supporting ADLs due to certain activities having set locations where they are performed – such as cooking and bathing. Some examples of these are Bluetooth and Ultrasound beacons that allow a receiver and transmitters to determine location [163]. This is achieved by measuring how long the signal takes to reach the receiver. Machine-vision systems can also be used to determine location by using techniques, such as background subtraction, to establish the person's location. Machine-vision also allows you to determine what activity the occupant is carrying out [164].

An increasing consumer trend that is also witnessing adoption within the domain of smart environments is the use of wearable technology [4]. Wearable technology offers new opportunities to AAL by enabling data to be gleaned not only from the environment but from the occupants of that environment [165]. Wearable sensors range from consumer activity monitoring devices produced by companies, such as Fitbit and Apple, which capture data relating to heart rate, sleep monitoring, and activity tracking. Towards complex cutting edge technology that, for example, embed sensors within fabric [166]. Beyond these, the emergence of head-mounted wearable technology in the last decade offered a new paradigm in wearable computing. These devices offer a first-person view of the environment, an eye-level display, along with on-board processing and communication capabilities [167, 168, 169, 31]. With real-time scene processing, such as object/facial/text recognition, allows the creation of supporting technology for the purposes of AAL [9]. One potential methodology which has shown potential is the use of fiducial markers within an environment when coupled to a wearable camera [170]. This method of indoor localisation involves a small set-up in which fiducial markers are placed within known locations within the environment. Cameras are then worn by the occupant and traditional image processing techniques are applied (such as feature point recognition). To allow the detection of the feature points within the FoV. There are then compared to the detected feature points to the known template of the fiducial marker. The use of fiducial markers allows for fast detection,

Table 2.3: Brief overview of commonly used sensors within AAL [3, 4, 5].

Sensor Type	Common Use
PIR	Person Localisation/Movement Detection
Ultrasound	Person Localisation
Bluetooth	Person Localisation/Object Information
WiFi	Person Localisation
Video	Person Localisation/Object Detection/Facial Recognition/Activity Recognition
RFID	Contact/Tag Information
Pressure	Chair/Bed/Contact
Contact	Door/Cupboard/Opening/Closing
Accelerometers/Gyroscopes	Activity Recognition/Movement Detection/Object Interaction
Audio	Activity Recognition/Fall Detection
Radar	Occupancy Sensing/Activity Recognition/Fall Detection

however, due to the nature of a wearable cameras can result in lost detections due to motion blur [49]. Wearable technology when combined with machine-vision techniques has been highlighted in previous works as offering potential technological developments within the domain of localisation and activity recognition within an AAL context [4, 136, 7, 49, 171].

2.7.1 Traditional Indoor Localistation Methods

This Section presents a summary of the current state-of-the-art of indoor localisation methods which do not leverage machine-vision approaches. A number of works are reviewed, which have a focus on applying contemporary technology to support occupant localisation within the domain of AAL. The selection criteria for the localisation methods reviewed were that the main focus had to be on the localisation of the occupant within an indoor home environment.

Rahal *et al.* implemented a system using anonymous dense sensor placement along with Bayesian filtering in order to determine occupant location [172]. The system was tested using a scenario of an occupant's daily routine. The routine was performed by 14 subjects, one at a time. The system showed a mean localisation accuracy of 0.85, as the authors note, however, the system is only capable of supporting a single occupant [172] within a fixed environment.

Okeyo *et al.* developed a dense sensor-based solution incorporating a Multi-Agent System (MAS) in order to provide services to occupants within smart homes [173]. A MAS consists of a group of agents which are able to interact with one another with the goal of achieving their design objectives. Sensors were placed on specific objects that the user would interact with which would then record the time and location associated with that sensor to build contextual information. While the overall results were high (1.00, 0.88, 0.88 for Precision, Recall, and Accuracy, respectively) it still suffers from the inherent problems that exist with dense sensor-based methods, such as multiple occupancy and the need for sensor interaction. Along with the problem of the cost of installation, both in terms of financial costs but also the personal cost of having the system installed in an occupant's home. Due to the time taken to perform the installation and the invasion of privacy as the equipment is installed in the occupant's own home can also add an additional burden onto the occupant and could act as a barrier to uptake.

Kanaris *et al.* [174] developed a system to provide indoor localisation through the use of BLE (Bluetooth Low Energy) devices and IEEE 802.11 Relative Singnal Strength Indicator (RSSI) fingerprinting. This was tested in an indoor environment of approximately 160m^2, six D-Link 802.11 Access Points were used to comprise the 802.11 wireless network while the BLE network consisted of four 802.15 Estimote devices. Each device was placed within a different room within the test environment. The information provided by the BLE and 802.11 was fused via the use of a novel i-KNN algorithm, resulting in significantly improved accuracy when compared to using 802.11 fingerprinting alone, accuracy was reduced from 4.05m to 2.33m.

Tariq *et al.* [175] developed a system to provide indoor localisation utilising capacitive sensors along with investigating various ML approaches to determine which algorithm offers the best performance in an indoor localisation scenario. To test the system four capacitive sensors were placed on the wall in a 9m^2 room, the data was labeled with the occupant's position in order to train the classifiers. A range of ML algorithms were tested to determine which would offer the best performance, with Random Forest offering the best results of accuracy, precision, and recall all exceeding 93% with an average error rate of 0.05m.

Belmonte-Fernández *et al.* [5] created a system to provide indoor localisation through the use of Wi-Fi fingerprinting coupled with a smart-watch to acquire the AP signal strength. The system was tested in three separate indoor environments ranging from 62m^2 to 120m^2. Four different datasets were gathered, two to train the system (each containing 50 samples for each location) and the remaining two (each containing 100 samples for each location) to validate the performance of the system. Results shown an average accuracy of 71.07% across all scenarios for all the experiments performed.

Antoniazzi *et al.* [176] created an indoor localisation system to locate occupants via RFID, the occupant was required to carry an RFID tag on their person that is detectable by readers throughout the environment. The readers transmit the coordinates of the detected occupant, reporting an error rate ranging between 12.08% – 21.79%.

Jiménez *et al* [177] combined the use of a smart floor, binary sensors, and RSSI received at a smartwatch from BLE beacons deployed within a smart environment, the smart floor device was regarded as the ground truth in order to estimate the location accuracy of the binary sensors combined with RSSI. The experiment took place over ten days with each of these days segmented into three distinct periods (morning, evening, and afternoon), the system accuracy over the ten day period demonstrated that the system was accurate to within 1.5m in 80% of cases.

Maghdid *et al.* [178] created a tracking system utilising smartphones, incorporating the on-board Wi-Fi and sensor devices, such as gyroscopes and accelerometers, to provide indoor localisation for occupants. Their approach used RSSI between the smartphone as the receiver and wireless access points within the environment combined with a Dead Reckoning (DR) measurement from the on-board sensors. This fusion of sensor data made use of an extended Kalman Filter which periodically compensted for the inaccuracies of the DR measurements via the use of the RSSI data. Their results show a positioning error within 2.5m.

Bianchi *et al.* [179] created an indoor localisation system utilising RSSI fingerprinting via a ZigBee wireless sensor network (IEEE 802.15.4), the occupant's location is also estimated from their interaction with devices within the environment. Each occupant within the environment was required to wear a MuSA (MUltiSensor Assistant) device which allows the occupant to be uniquely identified and provides their location through RSSI collected from ZigBee routers placed throughout the environment. They found they were able to achieve an accuracy of 98% within a home environment.

Kolakowski [180] developed a system utilising BLE combined with proximity sensors to provide a higher level of accuracy when compared to a BLE system. The system requires the occupant to wear a tag which continuously sends out BLE packets which are measured by receivers within the environment which measures the RSSI. The proximity sensors also perform independent location estimation. The measurements from the BLE and proximity sensors are transmitted to the system controller for the actual location to be determined by combining the BLE and proximity measurements. The system achieved a trajectory error rate of 0.27m with approximately 10% of the results having an error rate larger than 1m.

Sansano *et al.* [181] developed an indoor localisation system combining the use of Inertial Motion Units (IMU) within a smartwatch combined with Wi-Fi fingerprinting. Data was collected by four occupants within their personal homes for a period of two months, the occupants were asked to manually label intervals of time during which they were in a particular room performing ADLs with the system showing an F1 score of 0.92.

Vesa *et al.* [182] developed an indoor localisation system which utilised a smartphone coupled with Bluetooth beacons, they employed a ensemble based solution which combined a Multilayer Perceptron with Gradient Boosted Regression along with K Nearest Neighbours. The system was tested in a smart environment of $75m^2$ which was made up of four main rooms, their solution achieved an average localistaion error of 0.4m.

Kolakowski *et al.* [183] created an indoor localisation system which was comprised of BLE and UWB (UltraWideBand) nodes which were attached to the occupant and to various localised objects of interest within the environment. Anchor nodes were then placed within the environment to measure the UWB packets arrival time and measure the BLE signal strength to determine the location of a particular tag (either the occupant or an object of interest). The system was tested in an environment of approximately 79m^2. Eight anchors were placed within the environment being fixed to the walls or furniture close to the walls with at least one anchor was placed within each of the seven rooms. A user was asked to walk around the environment for ten minutes while wearing a tag as a lanyard. The system was able to correctly locate the tag in 95% of cases.

Bilbao-Jayo *et al.* [184] developed a system for indoor localistion leveraging a smartphone and smartwatch based on BLE technology. Bluetooth beacons were placed within each room, with multiple beacons being placed in larger rooms, such as the living room. The MAC address of each beacon was stored in a database along with its associated location within the environment. The smartphone/smartwatch devices were set to repeatedly scan for Bluetooth devices for ten seconds with a 15 second interval between scans. If a beacon was detected during the scanning window the approximate location of the occupant was determined by measuring the RSSI and TxPower values from the beacons. The system was tested in a home environment consisting of four rooms (kitchen, bathroom, bedroom, and living room) gathering a dataset which consisted of 267 location changes. The dataset was split into an 80/20 ratio for training and testing, achieving an accuracy of 67%.

Ceron *et al.* [185] presented a system for indoor localistaiton through the use of BLE beacons and an IMU which was located within the occupant's shoe. The system was evaluated within a pilot study consisting of 22 participants made up of 11 adults and 11 young people. The IMU device was set up to collect the occupant's acceleration and angular velocity while the BLE beacons were used to establish location via RSSI. The system reported a mean localisation error of 1.023m within the older cohort, and 0.986m within the younger cohort. The lack of a significant difference between the two cohorts within the pilot study suggested that the proposed method is effective across broad age ranges.

Parmar *et al.* [186] developed a system for indoor localisation through the method of voice fingerprinting from a single microphone array. The Seed ReSpeaker 6-mic circular array kit was utilised for data collection with the array placed centrally within an environment. Data was collected at 15 different training and testing locations within two scenarios collecting an occupant's voice from each location. Deep learning was used to train a Inception-ResNet-v2 model which resulted in localisation errors from the two scenarios of 1.56 and 1.48 metres.

From the review literature a number of limitations were found. The most common limitation is that of the need to install equipment throughout the environment [172, 173, 174, 175, 176, 177, 179, 180, 182, 183, 184, 186]. An additional limitation is that of the requirement for the occupant to wear a dedicated device, such as a tag or smartphone device [5, 176, 178, 179, 183, 181, 182, 183, 184, 185]. There are also identified challenges with regard to multiple occupancy with some of the systems [172, 173], particularly the dense sensing approach, due to anonymity of the data being collected by such a system. A final limitation found is that of required active sensor interaction to determine the location of the occupant [172, 173, 179]. As a result, the proposed apporach must offer a method of minimising equipment installation within the occupant's environment to reduce costs and the intrusiveness of the approach. Additionally, the requirement for the occupant to wear a device should be minimised in order to enhance a feeling of normality for the occupant. One potential solution to this limitation could be the use of smart glass as approximatly 74% of the adult population are required to wear corrective lenses [187]. An additional advantage would be the ability to identify which stream of data is related to which occupant.

2.7.2 Vision Based Indoor Localisation

This Section presents a summary of the current state-of-the-art solutions that facilitate indoor localisation utilising a machine-vision approach. A number of works are reviewed, which have a focus on applying contemporary technology using machine-vision techniques within the domain of AAL. The selection criteria for the machine-vision papers reviewed were that the main focus had to be on the localisation of the user within an indoor home environment via machine-vision methods.

Leotta and Mecalla [188] developed PLaTHEA (People Localization and Tracking for HomE Automation). PLaTHEA is a machine-vision based system that acquires a stereo video stream from two network attached cameras to provide support for AAL. Two cameras are placed in each room, working in stereo, in order to ensure that as much of the room is covered and that occlusions are reduced. Foreground extraction is then performed to determine if occupants are present in the scene. PLaTHEA also performs identity recognition using facial recognition. Facial recognition is performed using SIFT (Scale-Invariant Feature Transform) features from each face pose, which are then stored within a kd-tree data structure. At run-time, a Haar classifier [189] is applied to detect faces in the scene; when a face is detected SIFT features are extracted and compared to the saved features stored in the KD-tree for recognition [188]. There are, however, some potential limitations to the PLaTHEA system. Due to the system relying on static cameras it may not be possible to ensure that the entirety of the room is viewable or that occlusions may not occur due to the opening of doors, large furniture, *etc.* In addition, an issue that was identified by the authors, was when the system was monitoring a room with a wall greater than 10 metres then it was not possible to monitor without the use of costly acquisition hardware [188]. While the issue of cost is being addressed, there is also the additional cost of having to install multiple cameras within each room, that support is provided within. There is also the issue of multiple occupancy, due to the use of foreground extraction to identify occupants, while this is partially mitigated through the use of facial recognition, it also requires that all the occupants are known and have SIFT features saved within the system [188]. There is also the additional problem of the Haar classifier being reliant on the occupant's eyes being clearly viewed by the camera as this method of face detection will usually fail if the eyes are occluded [190].

Zeb *et al.* [191] developed a system that supported blind users, holding a web-cam, to navigating throughout a known environment. The web-cam continuously captured video frames from the environment, which were then processed for relevant markers. Whenever, a relevant marker was detected, the detection and identification module compared it to the stored markers in a database, returning a unique ID that associated the user's position and direction. While this system obtained a 98% success reate for detecting and identifying markers it required constant interaction from the user in the form of having to manipulate a handheld camera at all times, in order for the system to detect markers.

Rivera-Rubio *et al.* [192] developed a system that estimated the user's location through scene recognition. The experiment was carried out using an LG Google Nexus 4 and Google Glass. A dataset was gathered of the locations by recording a video of the occupant walking through the location ten times whilst wearing a recording device (50% split between the Nexus 4 and Google Glass). This included a combination of day/night acquisitions and occasional strong lighting from windows. The system was tested using multiple descriptor methods (three custom designed and three standard methods) following a standard bag-of-words, where low level features (such as colour) are extracted and applied to a visual analogue of a word, and kernel encoding pipeline, with HOG3D, a spatio-temporal descriptor, matching used as a baseline [192]. Results show errors as low as 1.6 metres over a 50-metre distance were achieved, however, for the purposes of AAL a greater level of refinement is required in order to distinguish where in a room the occupant is located and if possible what they are interacting with in order to provide relevant support. There is also the additional challenge of having to train the system to each environment that it is to be deployed within.

Zhang *et al.* [193] proposed a method of indoor location using still images captured at intervals from a smart-phone worn on a lanyard. This system had the goal of assisting those with impaired vision to navigate within an indoor environment. The system relies on collecting map data of a building, that describe features/descriptors along with their 3D co-ordinates, floor plans, and other location data. Images are then captured and sent at intervals from the smart-phone to a server for processing. Images are then matched against the template map of the building in order to determine location and offer directions should the user require them. Whilst this system works well for its intended use there are limitations when applied to an AAL situation. One problem, that the authors noted, was that there were null spots, were there was not enough features to create a map image, such as when the user makes a 90 turn, for example in a hallway or entering a room [193]. One other possible issue for an AAL application is that of intermittent image capture that may result in missing key information, such as a room transition or an interaction with an appliance, which could be vital for context.

Orrite *et al.* [194] developed a system entitled 'Memory Lane' with the goal of providing a contextualised life-blog for those with special needs. It chronologically tagged and ordered images and sounds perceived by the user to provide contextual meaning. A dataset of images of the occupant's environment was gathered and SIFT with RANSAC were applied to obtain feature points. During each RANSAC iteration a candidate fundamental matrix was calculated using the eight-point algorithm [195], normalising the problem to improve robustness to noise. Their system consisted of a wearable camera that systematically recorded still images as the occupant moved throughout the environment which would then be matched against the previously collected image dataset of the environment. A feature match correspondence was used to establish the distance of the occupant from the object. This involves generating a variable circle centred on the average position of the detected features and comparing it to the average position in the next image. If the radius increases, it can be determined that the occupant is moving closer to the object. Some limitations of this solution are the need to gather the dataset of the environment along with the inherent problems with intermittent image gathering.

Edwards *et al.* [196] created a fiducial marker system that concentrated on accuracy over run-time performance and compared the system against the ArUco marker detection algorithm [197]. The system was unique as instead of using corner or edge detection it used a radial sinusoid pattern which allows for a predictable appearance under perspective projection. The results indicated that, on average, pose estimation was twice as accurate than fiducial markers that rely on corner/edge detection, such as the ArUco system. This is achieved via a non-liner optimisation routine which estimates the fiducial marker's pose through minimising the difference between the predicted and actual appearance of the marker. The main limitation of this system is that the markers can not be identified individually and so need to be paired with a traditional fiducial marker system to provide an object/location identification.

Rituerto *et al.* [198] created a system that employed an Android phone, worn on a lanyard, using the ArUco algorithm to provide location/direction assistance to those with impaired vision. They created a digitised indoor map that stored information such as walls, corridors, room location, location of important sings/fiducial markers. The initial study was to determine the system feasibility of such a system. While the system was successful in providing direction to the occupant's it required them to steady the camera in order to return acceptable images, this would not be ideal in a real world situation due to the occupant's interacting with their environment in general daily activities.

Kapidis *et al.* [199] developed a system which utilised a wearable camera to determine location from key objects within the scene. They used the ADL dataset [164] which contains 20 videos of indoor activities with the Darknet framework [200] used to detect objects within a scene. A comparison was offered between CNN and LSTM based methods, with the CNN method resulting in an overall accuracy of 76% and the LSTM method offering an accuracy of 80%.

Domingo *et al.* [201] developed a system which combined a static RGB camera to determine the location of an occupant within an environment coupled with Wi-Fi fingerprinting to identify the occupant's identity once located via the RGB camera. An experiment was carried out using four RGB cameras installed within each room with the goal of locating 20 people moving freely throughout the environment. The system successfully located occupants in 79% of cases. Some issues were reported regarding occlusions and overlaps due to the static nature of the RGB cameras.

Martin-Gorostiza *et al.* [202] developed a system which combined a static camera with IR sensors with the goal of occupant localisation. The system consisted of a set of five IR receivers with a single fixed camera mounted on the ceiling. The system was able to locate the occupant with a precision of 1.5cm. Some of the reported issues include areas of the environment not being covered due to the fixed location of the camera.

Li *et al.* [203] developed a system to aid older occupants in keeping track the state of objects (*e.g.* oven *"on"* or *"off"*) and their past interactions with objects. The system utilised a wearable camera which was worn around the occupant's neck and would continuously detect fiducial markers in the FoV which were placed beside objects of interest. If a marker was detected then the camera would record a short video clip starting from when the marker was first detected and ends three seconds after the marker's last detection. The occupant can then review these video clips in order to ascertain the status of an object or to view their previous interactions with the object.

Hu *et al.* [204] developed a system to aid those with visual impairments navigate through an indoor environment through the use of fiducial markers. A panoramic ceiling view positioning framework which was based on a panoramic annular lens was used along with ArUco markers. The camera was head mounted on a helmet with a 180°FoV which included the ceiling and part of the walls and doors and did not include the ground within the FoV, with ArUco markers placed on the walls to define start and end points.

Kunhoth *et al.* [205] examined the performance and usability of two machine-vision based systems (CamNav and QRNav) along with a BLE system. CamNav utilises a trained deep learning model to recognise locations while QRNav makes use of QR codes as fiducial markers to determine the occupant's location. The systems were tested on ten blindfolded users who then had to navigate an indoor environment. The machine vision systems resulted in 30% less errors than the BLE system when providing users with real time assistance.

Quero *et al.* [206] developed a system to recognise daily objects within a smart environment using a wearable camera (GoPro Hero 5) with the goal of aiding in the collection large datasets. The occupant applies a bounding box to an object of interest to identify and label a static object. Background subtraction is then used to select the masked foreground object.

Buzzelli *et al.* [70] developed a system which involved the use of a single static camera placed within an environment with the goal of monitoring the elderly at home. The first stage in the system was to localise a person using a R-CNN (Regions with CNN features), to select the largest detected subject within the scene. DeepHAR (Deep Human Activity Recogntion) [207] was used to perform activity recognition. It infers the action through the explicit representation of the subject's inferred skeleton. They found that while the method offered high accuracy it faced challenges to multiple occupancy, in particular, the visitation of health care assistants.

Uygur *et al.* [208] developed a system to provide indoor localisation that uses input from a 360°camera to localise a user on a 2D map. They found the system was robust to partial blockage to the camera's FoV and did not require highly accurate maps. The sperical camera allows more data to be collected to attempt to overcome the issue of rooms mostly consisting of blank walls which are typically featureless. The features the system was designed to recognise were architecture features such as windows and doors, however, some issues were found such as doors being common within a large building and were not effective in reducing uncertainty in larger experiments. Other problems such as windows being difficult to detect due to their location within a wall and their close proximity to other windows.

Košt'ál and Slabý [209] presented a system which used novel fiducial markers to aid in localistation within spatial scenes. They tested their system using a total of 18 markers with five videos being recorded outside and thirteen videos being recorded inside an environment. The dataset containted 385 training images, with 110 validation images, and 55 test images. The ground truth bounding box on all images where manually tagged by a human expert. They found that testing with real world videos were crucial as it introduced motion blur that occur in natural camera movement, the results demonstrated a precision of 0.981 and a recall value of 0.927.

Li *et al.* [210] implemented a system which allows a user's location to be determined through a picture of the surrounding environment. An Android mobile phone (Lenove Phab 2 Pro) was used along with a depth camera (Intel RealSense D435) and the system was initially tested on the ICL-NUIM dataset which consists of RGB-D images from two indoor scenes – a living room and an office scene. The system was then tested in real world scenes within the BJTU lab space where a total of 144 images are collected, the resulting algorithm achieved an accuracy of 93%.

Zhou *et al.* [211] presented a system which leveraged machine-vision tools combined with a Convolutional Neural Network (CNN) to identify markers within complex scenes. The system was tested on the Pascal VOC dataset which contains approximately 30,000 images contained within 21 categories. Through testing it was shown that the system had a strong resistance to complex internal environments/background along with providing a high accuracy, in terms of positioning, and a fast processing speed.

Tabuchi and Hirotomi [212] developed a system utilising fiducial markers to aid those in cognitive decline to carry out the task of cooking. Fiducial markers were attached to objects of interest within the kitchen, such as cutting board, sink, stove, *etc.*. The system was tested as both a fiducial marker and a markerless sytems were the object recognition was used to detect the objects of interest. The results demonstrated that the system utilising fiducial markers operated approximately nine times faster and achieved a higher F-measure than the markerless system. An overall accuracy result of 70% was achieved by the system in a range of environmental settings.

From the reviewed literature a number of limitations were found. The most common limitation was that of training being required before the system can be applied to an environment, or moved to a new environment [70, 192, 194, 199, 205, 206, 208, 209, 210, 211]. An additional limitation that was found was that of occlusion within the video stream due to camera angles or objects blocking the FoV [70, 188, 193, 201, 202, 204, 208]. Intermittant image capture was also found to be a limitation, reducing the information that can be determined from the environment and potentially missing interactions [193, 194, 202]. A number of secondary limitations were also found, such as the issue of multiple occupancy [70, 188], required interaction by the occupant [191, 206], and the requirement to wear a device on a day to day basis increasing the burden on the user [198, 210, 204, 206]. The proposed approach must offer a method of removing the need to train the system to individual environments, allowing the approach to be applied to multiple environments without the need to retrain. Additionally the challenge of occlusion and intermittent image capture will need to be addressed to ensure that information, such as object interactions, are not missed.

2.7.3 Summary

From the various methods that have been reviewed it can be seen that a wearable camera offers many advantages over comparative systems, along with some advantages that are unique to a wearable camera, such as an egocentric view of the environment. This egocentric view helps reduce occlusions and also offers the ability to view which objects the occupant is interacting with. Additional advantages include being less sensitive from signal interference from other devices and minimal cost in terms of installation and long term maintenance [213]. Table 2.4 provides an overview of the advantages and disadvantages of the various localisation methods that have been reviewed in this section.

Table 2.4: Summary of advantages/disadvantages of sensor modalities [6, 7, 8].

Localisation Method	Advantages	Disadvantages
PIR	Low cost, low power	Intrusive installation, occlusions, multiple occupancy.
Bluetooth	Low cost	Limited range, installation, forgetting to wear device
Wi-Fi	Present in most homes, low cost, identify multiple occupants.	Multiple AP's for increased accuracy, forgetting to wear device
RFID	Low cost, low power	Interference, signal loss, forgetting to wear device.
UWB	Reduced interference from narrow-band/carrier waves, high accuracy.	High cost, metallic interference, forgetting to wear device
Dense Sensor Placement	Low cost, object interaction.	Intrusive installation, no occupant identification, requires occupant interaction.
Capacitive	Low cost, low power	Low coverage, electromagnetic interference
Static Video Camera	Occupant identification, object interaction.	Occlusion, privacy, intrusive installation, limited coverage.
Wearable Camera	Occupant identification, object interaction, no installation, reduced occlusion, improved coverage.	Privacy, forgetting to wear device.

2.7.4 Video Based Activity Recognition

This Section will present an overview of activity recognition solutions within the domain of AAL which leverage machine-vision based approaches. A number of works are reviewed, which have a focus on applying contemporary technology using machine-vision techniques within the domain of AAL to perform activity recognition. The selection criteria for the machine-vision papers reviewed were that the main focus had to be on activity detection of the user within an indoor home environment.

Giannakeris *et al.* [214] developed a system to perform activity recognition from a wearable camera, the ADL dataset [164] was used to test and evaluate their system. They used a Bag-of-Micro-Actions scheme using Gaussian Mixture Models (GMM) clustering with Fisher vector encoding to detect the activities. Their system achieved an accuracy of 57.14% on the ADL dataset.

Noor and Uddin [215] created a system to detect activities within an egocentric view using SIFT to detect feature points. The model was trained using an Artificial Neural Network (ANN), and leveraging Hidden Markov Models (HMM) to account for the various sequences that make up an activity. The system was tested on two datasets, the TUM Kitchen dataset [216] and the GTEA Gaze+ dataset [217]. The TUM dataset consisted of first and third person videos from five cameras with ten subjects, the GTEA dataset consists of a camera built into a pair of glasses with data being collected from ten subjects. The results show an accuracy of 96% on the TUM dataset, and an accuracy of 90% on the GTEA dataset.

Zuo *et al.* [218] developed a system to detect ADL within an egocentric view. Their system used an egocentric video stream from a Tobii Pro Glasses 2 [219] device which is then segmented into a set of video clips, each of which correspond to a specific activity the occupant carried out. An initial training dataset was gathered containing 50 interaction clips containing the following ADL: greeting, passing a ball, paying, shaking hands, and talking. These video clips are then classified as a particular ADL by applying a gaze-informed recognition approach. The system showed an accuracy of 97.32%.

Yu *et al.* [220] created a system to detect ADL through the use of an egocentric camera which supplies data via a photo stream and an IMU using an LSTM network. The system was tested on two datasets, eButton dataset [221] and the multimodal egocentric dataset established by Song *et al.* with Google Glass [222]. Their system achieved an average accuracy of 77% on the eButton dataset and on the multimodal dataset an average accuracy of 80% was achieved.

Diete and Stuckenschmidt [223] developed an egocentric multimodal based approach to activity recognition utilising smart-glasses and a chest mounted tablet. A pre-trained Neural Network which used the overlap of the subjects hand and the objects within the frame to determine object interaction was used. Two separate ADL datasets were used to validate their approach. The first dataset was gathered by the research team, the second was the CMU-MMAC dataset [224]. The system acheived an F1 measure of 79.6% on their gathered dataset and an F1 measure of 59.4% on the CMU-MMAC dataset.

Yu *et al.* [225] developed a system using a Kinect v2 sensor to recognise a range of 12 ADL such as, lie down, get up, comb hair, sweep the floor, *etc.* In order to attempt to address the issue of occlusion the sensor was mounted on the ceiling of the environment while being angled as to still offer a horizontal view plane. A dataset was collected from an elderly occupant living independently in their own home with activities being carried out naturally rather than in a prescribed manner. Their system achieved an average accuracy of 91.64%.

Massardi *et al.* [226] developed a system which used an Intel RealSense D-435 RGB-D camera mounted on a robot in order to detect activities of an occupant. They created a dataset of various ADL which included, making tea, making hot chocolate, and making coffee. The datasets were split into four different categories (category one, category two, category three, and category four), category one and three were used for training the system with categories two and four were used for testing. On average their system achieved an 80% accuracy.

Su *et al.* [227] developed a system which used a Deep Neural Network (DNN) with the aim of recognising ADL for supporting occupants aging independently at home. The list of activities that the system recognised was, standing, bending, squatting, sitting, eating, raising one hand, raising two hands, sitting plus drinking, standing plus drinking, falling. Data was collected containing all ten activities was gathered and manually labeled, the system showed an average accuracy rate of 95.1%.

From the reviewed literature the main limitation found was that of a need for the system to be trained for an environment, including a lengthy data collection phase [214, 215, 218, 220, 223, 225, 226, 227]. Additional limitations that were found included the issue of occlusion [225, 226] and the necessity to wear a device [223, 226].

2.7.5 Dealing with Uncertainty

This Section will present an overview of activity recognition solutions which implement evidential reasoning to aid with handling uncertainty found within real world data. Multiple techniques exist to deal with uncertainty found within data, an overview of some common methods are discussed below:

Baye's Theorem

Baye's Theorem [228] is a logical approach to revise the probability of a hypotheses being true when new evidence is supplied. Initial probabililty values within Bayes' Theorem are supplied from historical probabilities within existing data. Probabilities are then updated as new data becomes available allowing a method of revising existing predictions when given new or additional evidence [229]. Within the field of activity recognition Baye's theorem would use sensor data (such as vision, sound, or contact sensors) to estimate the probability that an activity has been carried out. As new sensor data is collected the probabilities of each activity being carried out is updated with the goal of attempting to determine what activity the occupant is carrying out in real-time.

Fuzzy Logic

Fuzzy logic can also be used to help with reasoning when uncertainty is present in the data [230]. One implementation of fuzzy logic is the the fuzzy Tsukamoto model, which is an alternative method of dealing with uncertainty by describing the relationship between the input and output via fuzzy "if-then" rules [231]. Fuzzy logic can be used to create a fuzzy inference system that evaluates the occupant's activity levels on a sliding scale rather than discrete categories. For example, if the system detects increased movement and social interactions then the system could determine that activity levels are higher than usual rather than simply stating active or inactive.

Monte Carlo

The Monte Carlo method has also been proposed as a way of dealing with uncertainty within data [232]. The Monte Carlo method is based on a mathematical model that determines the result based on random variables that can affect the outcome. This method is suited to estimating an outcome from the product of random variables, including sources of uncertainty [233]. In activity recognition the Monte Carlo technique can be applied to determine the most likely sequence of activities from sensor data. Typically the transitions between different activities are modeled and can use the sensor data to infer the most likely sequence of activities performed by the occupant.

Dempster-Shafer (DS) theory

DS theory has also been suggested to be a potential solution to dealing with uncertainty in data [234]. DS theory is an evidence theory framework for reasoning with uncertainty, the theory allows the combination of evidence from different sources to determine a degree of belief on the outcome [235]. DS theory offers a number of advantages over alternative reasoning techniques, such as the ability to combine various evidence types from various sources [236]. Additionally, traditional theories typically assign a probability to one possible event, however, in DS theory probabilities can be correlated to multiple possible events as well as offering the ability to represent the uncertainty of systems without further assumptions [233]. DS theory also offers additional advantages when applied in a multi-class problem [237]. This is due to DS theory applying a mass value to every possible class allowing the most likely class to be easily determined by comparing the mass values and selecting the class with the highest mass [237].

Due to these advantages DS theory is deemed to be the most suitable for this research, such as its ability to combine and manage conflicting sources of evidence. Additionally, DS theory allows for differing weights to be applied to different sources of evidence allowing for a larger weighting to be applied to markers that are more reliable or have a larger bearing on the likelihood of the activity being carried out. Lastly, DS theory is able to deal with situations were there may not be data available, such as through corruption or sensor failure.

A number of works were reviewed which has a focus on applying DS theory to the field of activity recognition. The selection criteria for the reviewed papers were that the main focus was on the application of DS theory to support activity recognition within a home environment.

Alcalá *et al.* [238] developed a system to monitor ADL behaviour through smart meter data on two datasets, the Household Survey dataset and the UK Domestic Appliance-Level Electricity dataset. DS theory was then implemented with the goal of detecting abnormal human behaviour within the environment. It was found that implementing DS theory was shown to be more sensitive to the pattern deviations of abnormal behaviour and less susceptible to false positives, in particular when there were long periods of inactivity. However, it was noted that this method was most suitable to carry out coarse monitoring of older occupants with a noticeable reduction in false alarms when compared to other methods due to DS theory's ability to handle peroid of inactivity which was modelled as an increasing uncertaintly.

Zhao and Li [239] implemented an abnormal activity recognition model based on an ontology and DS theory. An ontology was developed which included a number of basic activities (*e.g.* sit down, lie down, run, walk *etc.*), the location and duration of the event was then utilised in order to determine if an activity was normal or abnormal. For example, if an occupant was detected being within the kitchen and their activity was detected as lying down it could be inferred that the occupant may in some danger. DS theory was then used to handle the uncertainty within the sensor data to provide a more accurate estimation of the activity the occupant was carrying out.

Machot *et al.* [240] developed a system which uses DS theory within the domain of active and assisted living to support occupants carrying out their ADLs. Their implementation was tested upon the HBMS dataset of binary sensor data. The HBMS dataset contains five activities – watching TV, shopping, checking blood pressure, getting a drink, and preparing a meal. They achieved a 96.76% accuracy when testing on a subset of 10-day observations from the HBMS dataset. However, it was noted that the method had the disadvantage of requiring previously collected knowledge about the occupants' and the sensors.

Sfar and Bouzeghoub [241] presented a system for the detection of anomalous behaviour occurring within a smart home environment utilising DS theory. Their system was tested using the Hadaptic and opportunity dataset containing data from three participants carrying out three routines. The system was found to have an accuracy of 91% for the detection of abnormal behaviour when a suitable time window size was set. It was noted that when the time window fell below 180 seconds that DS theory became less efficient and was most efficient when the time window was proportional to the activities.

Venkatesh *et al.* [242] implemented a system for activity recognition within a smart environment using ML methods combined with DS theory to improve the overall recognition performance. Their approach was validated using a real world dataset from the UCI ML repository. It was found that combining a Probabilistic Neural Network (PNN) with DS theory was best in-class solution achieving 91.2% reliability for the detection of activities compared to an 85% accuracy when the DS component was not present. Bhowmilk and Mojumder [234] developed a system to monitor home and health parameters, such as temperature, pulse rate, SpO_2, *etc.*, with DS theory being utilised to aggregate data from multiple environmental sensor sources. The goal of the system was to estimate the threat level to occupants within a smart environment with DS theory was used to monitor four parameters, temperature, humidity, SpO_2, and carbon monoxide. It was shown that DS theory was capable of dealing with uncertainty in the data, in particularly missing data.

2.8 Challenges and Opportunities

Based on the review of the literature a number of challenges have been identified. These include the need for extensive training/fingerprinting for each unique environment. Multiple occupancy and unreliability in the data (*e.g.* sensor failure, interference, data corruption, false positives, *etc.*). This section will detail these challenges and present how this thesis aims to mitigate these challenges.

2.8.1 Cold Start

One of the problems that face many systems within the domain of is that of the "cold start" problem, where no data currently exists to train the system on [243]. This is a particular issue within data and context driven systems, which, will require a large amount of data surrounding the problem in order to learn to recognise locations or activities. As this data normally needs to be gathered before the system can be used in order to train the system to the environment. One example of the type of data that is required to be gathered is that of training data, where a large amount of labelled data is necessary. For example, should the model want to recognise the activity of cooking then a large amount of data of the occupant cooking will be required for training. Additionally, during the "cold start" phase any quality issues within the dataset, such as noise or missing data, can be more pronounced. In the cases of systems utilsiing RSSI, this phase will have to be carried out in each new environment due to layout changes within the environment, which will mean a new, unique set of "fingerprints" will need to be learned.

Systems using machine-vision techniques for object/scene recognition will also have to gather initial training data. While significant efforts have been made to produce datasets for this purpose [244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256], additional data will have to be gathered within the environment the system is to be deployed. This is due to differences between objects (*e.g.* different manufacturers of products), or in the case of system using natural fiducial markers within a scene, these will also need to be learned, as they will be unique to that environment [70].

Data augmentation has also been proposed as a potential solution to generating a dataset. Data augmentation involves applying various transformations to the existing dataset. This could include rotations, scaling, or adding noise in order to create new training samples [257]. It can be useful when working with imbalanced or small datasets; however, it should be noted there are disadvantages of data augmentation such as that the data quality could be affected due to the generation of unrealistic or irrelevant data. Additionally, as data augmentation can only generate variations of the existing data this would result in limited diversity within the dataset despite the increased dataset size [258]. As data augmentation cannot create new, original data no new features/information, which was not in the original dataset, would be generated.

This thesis investigates this problem using generic fiducial markers which can be placed throughout the environment. Each marker will have the ID of a particular object of interest which are common to the majority of environments. Examples include: kettle, TV, fridge, microwave, *etc.* This will mitigate the need for a training phase as the system will be pre-loaded with the suite of markers which can be applied in their relevant locations in any environment without the need to learn the new environment. Chapter Four explored the technical feasibility of applying fiducial markers for localisation within a live egocentric video stream. Chapter Five investigated and compared alternative approaches to fiducial marker design along with the feasibility of applying the system to multiple environments.

2.8.2 Multiple Occupancy

Most current AAL solutions assume the presence of only a single occupant within the environment [259, 172, 173, 175]. If there are multiple occupants that require support with ADL then it can be difficult to identify and offer appropriate support to the correct occupant.

Techniques that rely on RSSI have been a popular method to perform indoor localisation [174, 5, 176, 177, 178, 179, 180, 181, 182, 183]. While RSSI, when combined with a device such as a smartphone or smartwatch, can go some way to alleviate the issue of multiple occupancy (as each occupant has a device which can be assigned a unique ID). A common problem with multiple occupancy is when two or more occupants interact with objects within the environment within a narrow time window [260, 6]. If we consider an example where the oven door has been opened along with a cupboard door opened. There is insufficient evidence from the data to indicate which occupant has interacted with which object or even if this is indeed a case of multiple occupancy and both objects were coincidentally within reach at that point in time. This can lead onto further problems when attempting to learn behaviour as key components of this are when a specific occupant carries out a specific activity along with how they undertake that activity. This allows the detection of an occupant who is having particular trouble completing a specific ADL or if they are not undertaking the activity whatsoever. Due to the conflicts in sensor data, from having multiple occupants within the same environment, it can be difficult to separate these events into occupant specific events.

Possible techniques to mitigate this issue include occupant worn ID tags, such as RFID, that uniquely identify each occupant [261]. Video has also been used to identify the occupants within the environment [262], however, both these solutions introduce challenges of their own. Video from fixed cameras can suffer from challenges such as occlusion where the occupant of interest may not be visible due to objects blocking the camera's FoV. Typically solutions relying on RSSI are not able to provide a fine enough accuracy to reliably distinguish between multiple occupants who may be in close proximity to each other [263, 54]. This can be further compounded as RSSI methods can be affected by signal interference, layout of the environment, *etc.* [263, 54, 6]. These issues can be further compounded when the occupant has visitors or carers that may call in on a regular basis, as the visitors will not be recognised by the system but the occupants may still require support.

This thesis aims to mitigate the multiple occupancy challenge by the use of a wearable egocentric camera which will provide a first-person view of occupants within the environment. This will allow indoor localisation to be applied to each occupant's unique view point regardless of how many other occupants there may be within the environment at that time. As each occupant will have a unique video feed each marker detection will be associated to the an individual occupant and thus support can be targetted towards that particular occupant. The use of an egocentric view will also enable the occupant-object interactions to be collected allowing activity recognition to be carried out along with the indoor localisation of the occupant.

2.8.3 Unreliability

As AAL systems are reliant on collecting data from the environment, a common problem these system will face, will be that of sensor unreliability [175, 181, 226]. Unreliability can take various forms, the simplest of which is that of sensor failure where sensors may report FP, such as reporting a sensor event when none exists or conversely failing to report a sensor event when one did take place. This is known as a False Negative (FN).

Unreliability can also be introduced through “noise”. This can take various forms depending on the type of sensors used. In the case of RSSI based systems this can be caused by signal interference, either through other devices utilising the same frequencies or simply through signal degradation due to passing through solid objects [7]. In the case of a system utilising vision sensor, noise can take the form of extreme variation in brightness/colour in images, extreme levels of motion blur which may render the frames unusable along with issues related to wearable cameras, such as auto-focus [264]. Additionally, the use of intermittent image capture can also introduce uncertainty into the data due to missed events. Additional forms of sensor unreliability can include inaccurate environmental sensors, this can be caused by sensor drift or through sensor malfunctions. This can be mitigated through the regular calibration and maintenanice of environmental sensors along with data fusion techniques from multiple sensors in an attempt to improve accuracy.

A large number of the reviewed systems all reported issues with having to deal with noise/missing data/incorrect sensor events [172, 176, 178, 180, 182, 194, 196, 202, 204, 70, 208, 218, 223, 225, 227, 177, 188]. Chapter Six details an approach for taking account for unreliability within sensor data by utilising Dempster-Shafer theory for reasoning with uncertainty in the data stream. Dempster-Shafer theory is a framework used to handle uncertainty when there is missing/conflicting information, it allows you to combine evidence from different sources to take account of uncertainty within the data.

Chapter Four details the approach for performing indoor localisation on an egocentric live video stream, utilising machine-vision techniques. This approach leverages wearable technology, Google Glass, to facilitate a unique first-person view of the occupant’s immediate environment. Machine-vision techniques are employed to determine an occupant’s location via environmental object detection. This method provides additional secondary benefits such as first person tracking within the environment and lack of required sensor interaction to determine occupant location.

Chapter Five also considers distance estimation in order to aid in filtering out false interactions. A linear filtering method is applied along with a fuzzy membership function to estimate the degree of occupant interaction, to assist in removing FP generated by the occupant.

2.9 Conclusion

This chapter has provided an literature review of the technologies and applications involved within the domain of AAL. It has introduced the concept of ADL along with the concept of a smart environment and how technology can be applied to provide support in order to improve QoL for those that would normally require full time care or institutionalisation. An overview of the technology used has also been presented along with an overview of the major techniques that can be applied to AmI, namely, data, knowledge, and context driven approaches. Current methods of supporting ADL suffer from common problems such as the cost of retro fitting an environment along with the intrusiveness such an installation will incur. Other problems exist with current support of ADL such as the "cold start" problem where a large amount of data needs to be collected for pattern recognition through data mining. When the system is initially installed there is no data to be processed, therefore support will not be available.

Key challenges that AAL faces have also been presented and discussed, such as those of multiple occupancy, training, occlusion, and unreliability in the data. Future challenges include the personalisation of support to each individual occupant's needs, each occupant's condition will deteriorate at differing rates therefore support will need to be tailored to the individual. These challenges will need to be addressed in the future, if the vision of AAL is to be achieved in a real world setting.

Chapter 3: Generation of Egocentric Datasets for ADL Research

3.1 Introduction

Chapter 2 presented the current state-of-the-art in support of ADL within the home. This Chapter details the approach taken to generate appropriate datasets for use within this thesis and has been made available to the wider research area [1]. The Chapter includes the design of the activities being recorded and the routines, along with an overview of the sensor technology that was used and the differing environments that were used to gather a more suitable dataset.

This Chapter will discuss the generation of a dataset using multiple sensor types within multiple environments, through both real world experiments and through the use of a simulation tool to generate datasets. While many research groups are sharing their activity datasets [265, 266, 267], due to the nature of human activity a diversity in experiment set ups are required in order to attempt to gather comprehensive datasets. This is also further compounded by the nature of sensor technology constantly evolving, which, requires additional datasets to be gathered to take account of the introduction of new technologies, as such no single dataset exists which is considered adequate [70].

3.2 Routines to Simulate Activities of Daily Living

As this research utilised smart glasses (Google Glass), a current dataset did not exist containing first person video footage of a range of ADL being carried out within a home that could be used. A brief review of common datasets are offered in Table 3.1; only datasets which included ADL were included.

As can be seen none of the currently available datasets are suitable, some of

[1] https://github.com/cshewell747/VisionData

Table 3.1: The number of activity classes, instances, and sensor type used in publicly available ADL datasets.

Dataset	Classes	Instances	Camera
IXMAS [250]	15	396	Fixed RGB
Hollywood 2 [252]	12	1,707	Film clips
ADL [164]	18	10 hours	Chest mounted GoPro
MSR [253]	16	320	Fixed Kinect
N-UCLA [268]	10	1,475	Multiple fixed Kinects
UWA3D II [269]	30	1,075	Multiple fixed Kinects
Kinetics [247]	400	306,245	YouTube video clips
DALY [256]	10	3,600	YouTube video clips
Charades [248]	157	9,848	RGB mobile phone
NTU [249]	60	56,880	Fixed Kinect

which are targeted towards a specific scenario while others are generalized along with large variations in data quality and consistency. None of these datasets are suitable for simulating an elderly occupant performing ADL within an indoor context. In order to overcome this limitation a dataset consisting of first-person footage, along with additional sensor data to aid in comparing the effectiveness of the machine-vision platform, was generated using the Google Glass platform.

A protocol was designed that was comprised of a range of activities to be carried out which were representative of daily routines, the protocol was carried out by a single researcher. With the goal of recognising the component locations (*e.g.* drinking water consists of kitchen door, glass cupboard, and sink) within each activity along with further investigation to determine if the activity could be determined via the component locations. If for example prepare/drink water is taken as an example activity, then the component locations would be the kitchen door, the cup cupboard, the tap, and then finally the kitchen door again.

The routines were derived from commonly performed household activities which consist of basic ADLs, such as ambulating and prepareing food [270]. The activities that make up each of the three routines were selected in a psudo-random method, ensuring that a range of activities which are representative of a real world routine.

In order to provide a variation in the simulated daily routines three routines were created. The first contained ten activities and the remaining two contained eleven activities. The activities ranged from simple activities such as drinking a glass of water to more complex activities, such as preparing hot food. The activities considered are presented in Table 3.2, with the full routines presented in Table 3.3. The routines were generated in an attempt to simulate a wide range of activities one might encounter within their daily routine, comprising of “basic” ADL, such as making a meal, to “insturmental” ADL, such as washing dishes and telelphone communication [270]. A variation in the number of activities that were carried out within each routine were also introduced to simulate people’s routines varying on a day to day basis.

Due to issues traditionally faced when migrating a AAL support system from one environment to a new environment multiple test locations were used in order to test the robustness of the system. Issues can include the need to retrain the system to the new environment, along with the intrusiveness of the install within an occupant’s home along with the associated finanical cost. The test locations were the smart lab within the Pervasive Computing Research Centre (PCRC) lab at Ulster University [271] and the smart lab within the Ambient intelligence lab at the University of Jaén [272]. These routines were performed under the same lighting conditions in order to minimise any potential discrepancy between identical activities in differing routines. The same routines were then carried out with the UJAmI lab but under varying lighting conditions to judge the effect lighting has on the effectiveness of the system.

In order to be able to confidently label the events and time stamps of the machine vision and binary sensor location systems, the ground truth was obtained from a time stamped video which provided a recording of the environment. The occupant’s location reported from the location systems were then compared to the ground truth from the video. Each routine was carried out by a single researcher in a structured manner to ensure repeatability. Table 3.3 details the number of activities within each routine.

It should be noted that there can be limitations when a single researcher is performing all of the data collection. One such issue is that of limited objectivity which can result in a subjective bias being introduced during data collection, this can be due to issues such as the individual intrepretation of the protocol. Additional issues that can be faced is that of reduced data diversity as a single researcher may collect data in a rigid manner which may result in less diversity in the final dataset. For example, activating sensors in a strict order when carrying out activities that would normally have a level of variation when carried out nat-

Table 3.2: The full list of activities that were performed during the three routines.

Activity Number	Activity
1	Prepare/drink water
2	Prepare/drink tea
3	Prepare/drink hot chocolate
4	Prepare/drink milk
5	Make/receive phone call
6	Prepare/eat cold meal
7	Prepare/eat hot meal
8	Watch TV
9	Wash dishes

urally, such as making a cup of tea. Lastly, having a single researcher collect all the datasets can result in a smaller dataset simply due to time restrictions when collecting the data.

3.3 Hardware Used for Data Collection

This Section will detail the hardware used to record the data, including the wearable sensor platform and lighting controls that were used when gathering the datasets within the labs at both recording sites; PCRC and UJAmI.

3.3.1 Binary Contact Sensors – TyneTec

Binary Contact sensors [273], refer to Figure 3.1, were used to provide a benchmark method to assess the viability of the machine-vision method to determine occupant location. Specifically, TyneTec binary contact sensors comprise a two-part magnetic based sensor. One part is a magnet and the other part is the sensor itself, the sensor is triggered when these two sections are separated. The sensor activates when the magnet is taken within or without range of the sensor sending a signal to the receiver which logs the event in a database. The data collected from the binary contact sensors will be used to provide a gold standard to assess the performance of the machine-vision system which is discussed in Chapter 4.

Table 3.3: A breakdown of activities that took place in each routine, along with the corresponding activity number.

Routine 1 (R1)	Routine 2 (R2)	Routine 3 (R3)
3	4	3
1	6	1
7	1	5
9	5	7
8	1	1
1	2	8
8	8	2
6	7	8
9	9	6
1	8	9
N/A	1	4

Figure 3.1: TyneTec ZXT434 Binary Contact Sensor.

The use of binary contact sensors offers a range of benefits over alternative technology solutions. One benefit is that of simplicity, binary contact sensors consist of a simple on/off switch that can detect whether contact is present or has been broken. This simplicity can make binary contact sensors a cost effective solution which can be simple to install within an environment, requiring a low amount of retrofitting to the environemnt. Binary contact sensors also offer high accuracy for detecting the presence of an occupant via the activiation of a sensor, such as when opening a door or cupboard. They also offer an advantage in terms of privacy as binary contact sensors do not transmit personal or identifiable data. Only transmitting that they have been activated along with a timestamp.

However, there also a range of limitations when utilising binary contact sensors. Due to the nature of binary contact sensors they can only detect two states – on and off. This limited level of granularity can make it a challenge to capture finer details and requires the occupant to interact with a sensor to collect any information. They are also limited when it comes to multiple occupancy as they cannot identify between different occupants, simply reporting that the sensor has been activiated. This can be a challenge when supporting older adults who may require assistance from family members or caring staff.

3.3.2 Wearable Camera – Google Glass

This research employed the Google Glass Explorer Edition [29]. It provides a first-person video camera, in addition to a full sensor suite of accelerometer and gyroscope, GPS; Table 3.4 provides a full list of the available sensors within Google Glass. User input can be gathered either through the touch interface or the natural language commands.

Table 3.4: A breakdown of Google Glass specifications.

Component	Specification
Operating System	Andriod 4.4
Display	Himax HX7309 LCoS 640x360
Camera	1280x720
Wi-Fi	802.11b/g
Bluetooth	4.0
Storage	16GB (12GB Available)
CPU	OMAP 4430 SoC 1.2Ghz Dual Core (ARM v7)
RAM	1GB
Sensors	3 Axis Gyroscope/Accelerometer/Magnetometer
Audio	Bone Conduction Transducer
Battery	570mAh 2.1V (7560 Joule)

The on-board processing capabilities of Google Glass consists of 682MB usable RAM (1 GB total – 342 MB reserved), and a dual core TI OMAP 4430 1Gz processor. The CPU can be set to four frequencies – 300Mhz, 600Mhz, 800MHz, and 1GHz. At high temperatures the Glass firmware limits the CPU to 600Mhz or 300MHz in order to cool down via power reduction [274]. The proposed method outlined in this thesis can be applied to any first-person camera, whether this is provided by an off-the-shelf solution such as Google Glass or a device as simple as a webcam. In the presented work Google Glass was streaming live video at a rate of 20 FPS at a resolution of 640x480.

The use of an off-the-shelf solution offers many advantages, firstly Google Glass are designed to be lightweight and ergonomic along with accepting prescription lenses. Due to the reliance on the occupant wearing the glasses at all times it is important that the device is comfortable to wear and by accepting prescription lenses it removes the need to rely on the occupant to remember to use the device when required. Google Glass also offers a user friendly interface allowing commands to be run via voice commands or touchpad controls which is an advantage given the typically lower levels of technological literacy among the cohort. However, there are some limitations when using Google Glass within the scope of this research. The main limitation of Google Glass, and that of other smart glasses, is that of battery life, due to the need to continuously run the camera alongside the small form factor of smart glasses results in a reduced battery life. Battery life can be further extended with external battery packs, however, with the current rate of advance in battery technology the battery life of future generations of wearable devices will be less of a challenge. An additional limitation is that of device cooling, which is achieved by reducing the clock speed of the CPU. At high temperatures, the Glass firmware limits of the CPU to 600Mhz or 300Mhz to cool down via power reduction which can result in reduced performance from the device.

The data gathered from the Google Glass device was used to determine the performance of the Glass device. This was done through comparing the accuracy of the locations detected by the Glass device with the accuracy of a dense sensing solution that was placed within the environment, which consisted of TyneTec sensors placed on objects of interest as detailed in Chapter 4. The data was also used to provide a comparision of the accuracy of other fiducial marker detection algorithms as detailed in Chatper 5. Additionally, the data was used in a later study to perform activity recognition within an environment to determine what activity an occupant was carrying out and if the use of probability theory could further improve the accuracy of the system as discussed in Chatper 6.

3.3.3 Lighting Detection – Sun SPOT

Sun SPOT (Small Programmable Object Technology) sensors [275], refer to Figure 3.2, were developed by Oracle to allow the development of new applications and devices. They consist of an embedded microprocessor running Java and offer a range of technologies such as IEEE 802.15.4 communication, built-in Lithium Ion battery, built-in ECC public key cryptography, and a range of built-in sensors. The key Sun SPOT sensor in the scope of this thesis is that of ambient light detection (wavelength measured in nanometers) to allow the assessment of how ambient lighting effects the accuracy of machine-vision systems.

Figure 3.2: Oracle Sun SPOT UDM3011 sensor.

3.3.4 Lighting Control

In order to facilitate the control of the lighting within the labs, roof mounted fluorescent lighting was used in combination with natural light control through the use of window blinds. The fluorescent lighting in each room consisted of a series of Philips TL5 HE 835 28-watt bulbs producing 86lm/W [276]. Table 3.5 details the lighting details within each room in the environment.

Table 3.5: Details of the lighting in the kitchen and living room section of the PCRC smart environment.

Room	Number of Bulbs	Total Watts	Total Lumens
Kitchen	3	84 Watts	7,224
Living Room	6	168 Watts	14,448

This allowed a consistent level of light to be controlled throughout the day while the protocol was being carried out within the PCRC lab. The same method of controlling the light using fluorescent lighting along with window blinds were used with the UJAmI lab to allow differing levels of lighting to be set in order to assess how this affected the performance of the proposed system. In order to create a "high" level of lighting (approx. 500 lux) for the experiment, all of the fluorescent lights were turned on and the window blinds were left fully open. Experiments were conducted at approximately the same time of day over multiple days in order to help control the amount of ambient light due to the time of day. To simulate a "medium" level of lighting (approx. 300 lux) the window blinds were closed and the fluorescent lighting remained on, and to simulate "low" levels of lighting (approx. 150 lux) both the window blinds were closed, and the fluorescent lighting were turned off.

It should be noted that there are some potential limitations of the lighting control. Due to the exact lighting level at each object of interest not being measured with only an approximate reading being taken it can be difficult to know if the lighting over a particular object of interest was lower or higher than expected. This could result in a higher or lower accuracy than expected. Additionally, the reliance on natural light as a component can make it difficult to ensure consistancy across all experiments. However, an advantage of this setup is that it allows a more accurate replication of a real world environment which would consist of natural lighting alongside ceiling lighting.

3.4 Deployment

Significant efforts [272, 271, 277, 278] have been focused upon establishing smart environments which allow the development and testing of emerging technologies along with the generation of datasets. These can be reduced to three main categories[271]:

Lab Environments – These are mainly research based environments and are typically located within research, academic, and industry locations.

Smart Environments – These are living environments which have been created for the sole purpose of evaluation and demonstrating newly available technology.

Smart Living Environments – These are living environments which have been designed to meet the real world living needs of people within their own home, typically a long-term implementation.

Any institution that is involved with the development or testing of smart environment technologies would have established some form of lab environment in which to perform testing or development of their systems. Lab environments vary in size and scope, from simply portioning off a section of a lab to the development and installation of a fully dedicated smart space. There are a number of notable examples of smart environments created for evaluation and demonstrating purposes. These allow the development and evaluation of various forms of technology in order to direct future research. There are notable research groups that have created these dedicated smart environments, such as the MavHome project [279], Gator-Tech [277], Aware Home [280], PCRC lab [271]. UJAmI Smart Lab [272], and the H^2AI – Human Health and Activity Laboratory [281].

The final categorisation is that of real world homes which will have occupants living within them on a day-to-day basis, this may require an external care provider depending on the level of care that the occupant requires and if this can be solely met by the smart environment. These homes are equipped with various levels of supporting technologies such as fall detection to fully autonomous homes that operate doors and windows among others [282], such as those developed by the University of Zurich [283].

Each of the categories offers a range of benefits for the development and testing of new technologies. Lab environments offer controlled conditions to allow for individual variables to be isolated and controlled to measure their effect on the outcome. This also allows for reproducibility to allow for the modification of individual variables and also to ensure that results are repeatable and that the technology operates in a consistant manner. Smart environments allow for the simulation of a real world setting to provide a more realistic context to test the technology within, such as performing activities within a home. Smart environments also offer the oppourtunity to observe how users interact with the technology in a controlled environment along with allowing the testing of interoperability between devices or technologies. Lastly, smart living environments offer the oppourtunity to perform real world testing/validation with a target user group. Additionally, it allows for data collection within a real world setting which can aid in further development or for training/updating new models.

3.4.1 Ulster University Pervasive Computing Research Centre Lab

The Smart Environment lab at Ulster University was established in 2009 by the PCRC, Figure 3.3 shows the layout of the PCRC lab along with the placement of the various sensors that are available throughout the environment. The lab in

Ulster offers binary contact sensors. The list below details the sensor type and their location within the PCRC lab. A floor pan can be seen in Figure 3.3.

- Binary Contact Sensors - 11
  - D01 – Kitchen door
  - D02 – Cup/glass cupboard
  - D03 – Tap
  - D04 – Tea/Hot Chocolate cupboard
  - D05 – Kettle
  - D06 – Fridge
  - D07 – Microwave
  - D08 – Cutlery cupboard
  - D09 – Living room door
  - D10 – Plate cupboard
  - D11 – Chair
  - D12 – Sofa
  - T02 – Phone
  - TV0 – TV

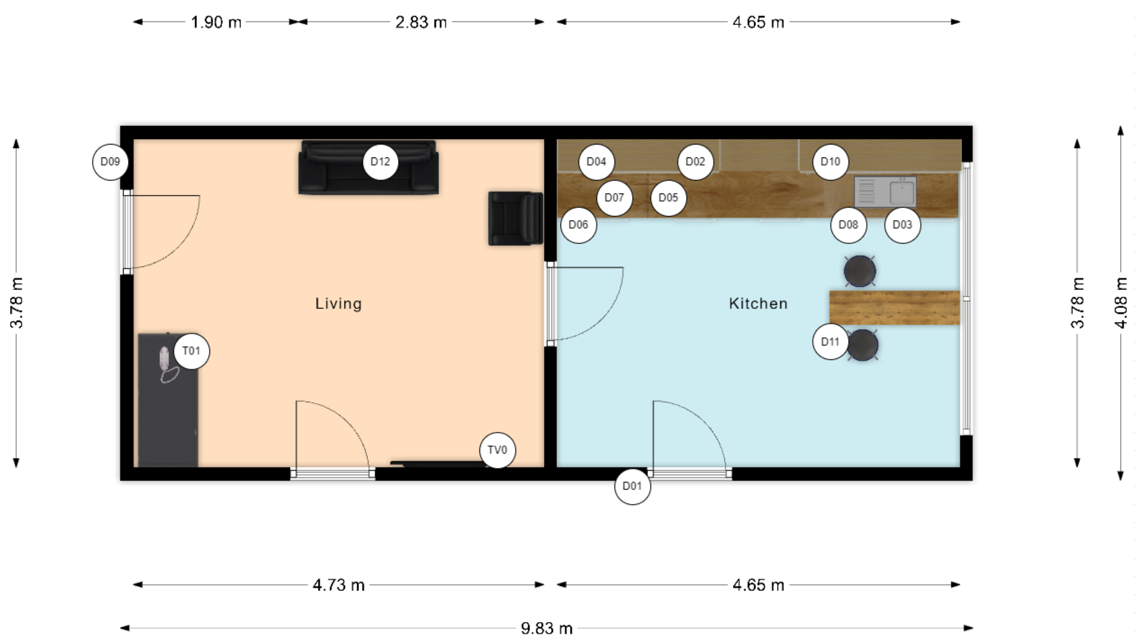


Figure 3.3: PCRC smart environment floor plan with sensor locations.

3.4.2 University of Jaén Ambient Intelligence Lab

The Smart Environment lab at the University of Jaén was created in 2018 by the Advanced Studies Centre in Information and Communication Technologies and Engineering (CEATIC). Figure 3.5 shows the layout of the UJAmI Smart Environment lab along with the placement of the various sensors that are available throughout the environment. The UJAmI lab offers binary contact sensors, Passive InfraRed (PIR), and Sun SPOT sensors. Images of the environment can been seen in Figures 3.6a and 3.6b.

(a) Living room view of the Jaén smart lab.

(b) Kitchen view of the UJAmI smart lab.

Figure 3.4: Images of the living room and kitchen within the UJAmI lab.

The list below details the all the sensor types and their location within Figure 3.5:

- Binary Contact Sensors – 12
  - D01 – Fridge
  - D02 – Microwave

- D03 – Wardrobe
- D04 – Dishwasher
- D05 – Plate cupboard
- D07 – Toilet
- D08 – Groceries cupboard
- D10 – Cup/glass cupboard
- K01 – Kettle
- TV0 – TV
- M01 – Front door
- T02 – Telephone

• PIR Sensors – 4

- SM2 – Bed
- SM4 – Bedroom door
- SM5 – Sofa
- WT0 – Tap

• Sun SPOT Sensors – 6

- SP1 – Above microwave
- SP2 – Above PC desk
- SP3 – Above living room cupboard
- SP5 – Above cooker
- SP6 – Above bed
- SP8 – Bathroom sink

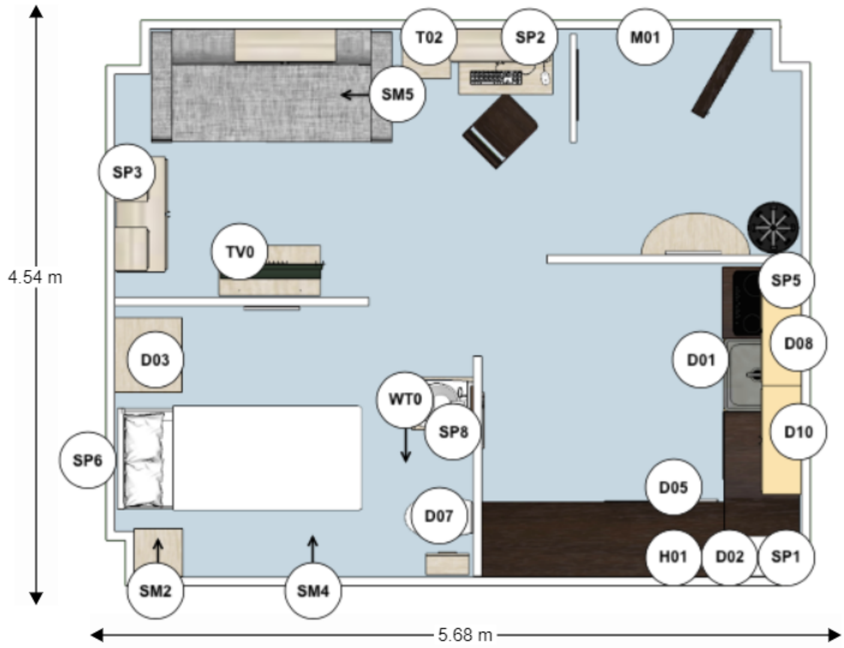
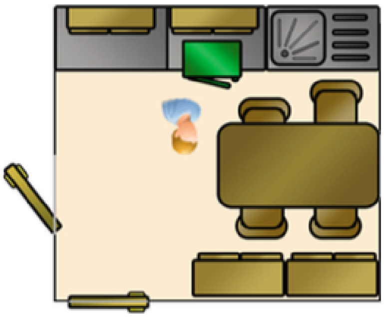


Figure 3.5: Sensor positions within the UJAmI lab, PIR orientation is shown by a small arrow.

3.4.3 Simulated Dataset – IESim

Due to the inherent difficultly of gathering large datasets within a smart environment a simulated dataset was created in order to provide a suitable large dataset which would allow the testing of technology and methods. Intelligent Environment Simulator (IESim) was a tool developed within Ulster University for the simulation of smart environments and sensor platforms [284]. IESim was designed to aid in the rapid creation of a simulation of a smart environment which could be populated with sensors and objects. It provides an interactive visual approach to allow its use by both technical and non-technical users to create novel environments in order to perform initial testing. Each routine from Table 3.3 was simulated twice within IESim simulating binary contact sensor data from a single occupant. Figure 3.6a and Figure 3.6b presents an example of an environment developed with the use of IESim. This simulated environment is designed to simulate the PCRC smart lab.



(a) Smart kitchen within IESim.

(b) The smart kitchen in the PCRC lab.

Figure 3.6: Images of the living room and kitchen within the PCRC smart Lab.

When compared to the datasets collected by a researcher the simulated datasets does not contain any missed sensor events and the sensor events are carried out in a strict order. Due to these issues it was felt that the simulated dataset did not represent the complexity of a real world scenario were variations may exist within the data due to the order of sensor activations. Additionally, in a real world scenario there may be missing or corrupt data, this can be due to many factors such as hardware/battery failure or interference. This simulated dataset could potentially be used within future work, one interesting avenue of investigation is that of utilising the simulated dataset to augment the real world dataset to increase the size of the dataset for training ML models.

3.5 Datasets Collected

A single participant generated the data in both the PCRC and UJAmI labs. Table 3.3 presents the three routines that were carried out at each lab along with the corresponding activities (experiment protocol available in Appendix A.). The three routines contained 175 sensor events in total, resulting in a total of 350 vision events over both labs. A simulated dataset was also generated through IESim, which produced a simulated TyneTec dataset of 651 sensor events. In total four main datasets were gathered:

1. A simulated data that was generated via the use of IESim, to provide a large dataset of synthetic contact sensor events.

2. A gathered dataset from the PCRC lab which provides a dataset comprised of egocentric video data from a wearable camera under consistent lighting conditions. This was coupled with binary contact sensor data to act as

a ground truth, this was used to determine the technical feasibility of the system.

3. A gathered dataset from the UJAmI lab which provides a dataset comprised of egocentric video data from a wearable camera under consistent lighting conditions. This was used to determine if it could be easily applied to multiple environments.

4. The gathered dataset from PCRC lab which provides a dataset of egocentric recordings from a wearable camera under varying lighting conditions with ArUco and custom fiducial markers to compare systems.

This provides a wide range of data on which to test the system allowing a method of comparing the ease of set up and installation within differing environments. Along with a wide range of data, complete with varying lighting conditions, on which to tell the reliability of the system. Further breakdown of the datasets are presented in Table 3.6.

Table 3.6: The full list of activities that were performed during the three routines.

Dataset ID	Classes	Instances	Sensor
1	9	651	Simulated binary contact
2	9	175	Google Glass
2	9	175	TyneTec ZXT434 Binary Contact
3	9	175	Google Glass
4	3	38	Google Glass

3.6 Summary

This Chapter presented an overview of current, commonly available datasets for ADL, and produced a novel dataset for indoor localisation and detecting ADL which have been made publicly available [2]. The datasets consisted of an occupant carrying out ADL while wearing an egocentric camera, along with binary contact data via TyneTec sensors. The datasets discussed within this Chapter offer advantages over other publically available datasets, such as those presented in Table

[2]https://github.com/cshewell747/VisionData

3.1, mainly through the use of fiducial markers applied to an environment which is viewed through a first person context. Additionally, the data recorded through the use of a head mounted wearable camera which offers a unique perspective of the environment along with helping to reduce issues such as occulusion. However, it should be noted that there are limitations to the collected dataset. Firstly, the small size of the dataset is a limitation when discussing training ML models which can require large amounts of data and can result in overfitting. Additionally, the use of custom fiducial markers restricts what algorithms can be used. As a large number of fiducial marker detection algorithms rely on pre-designed markers and typically cannot be modified to accept custom markers. [197]. Chapter 4 presents use of these datasets to propose a novel form on indoor localisation utilising a wearable camera and a mechanism to identify “key” objects within the environment.

Chapter 4: Indoor Localisation through Fiducial Marker Detection on Near Real-Time Wearable Video

4.1 Introduction

This Chapter presents a novel, in terms of technology used and method of localisation, solution to the challenge of occupant localisation within an environment. The work reported in this chapter has been published in [264] and [285]. The proposed method leverages smart glasses (Google Glass) and fiducial markers placed on key objects within the environment to determine location. This is achieved via the live streaming of a video feed from the front facing camera on the Glass device. The video feed is then processed and any fiducial markers within the stream identified. Each marker will have an associated I.D. which details the approximate location of the occupant. The novelty of this system is the use of a near real-time video stream to perform localisation through the use of fiducial markers placed on "key" objects within the environment via a smart glass device.

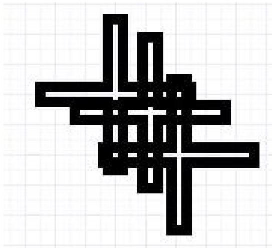
The main objectives of this Chapter were to present a review of the current state of the art of machine-vision based solutions that facilitate indoor localisation, to establish an experimental protocol to assess the viability in applying an indoor localisation system utilising a wearable camera, along with its feasibility to be applied to multiple environments. The results were validated at multiple locations (PCRC and UJAmI labs). Furthermore, a comparison of how the costs of the presented system compares against the costs of alternative sensor platforms for occupant localisation is presented as the finanical cost of such a system will be key to widespread adoption. The hypothesis considered that the use of a single wearable camera allows occupant tracking within an environment.

In order to assess the feasibility of the method, a protocol, as discussed in Section 3.2 was established to compare the presented method against an established method of indoor localisation; dense sensor placement [286]. A series of activities were performed within an environment and the location was recorded by both a machine-vision localisation system and dense sensor placement. To verify that the method was applicable to multiple environments (differing environmental layouts, lighting conditions, ease of installation, *etc.*) the experiment was recreated within two separate smart labs, the labs in the PCRC lab [287] and the UJAmI lab [272]. This Chapter details the rationale, architecture, methodology, testing, and evaluation of a system to facilitate indoor localisation through the use of a single "always-on" egocentric camera, implemented using the Google Glass platform.

4.2 Methodology

This Section proposes a solution to facilitate indoor localisation through the use of a single "always-on" egocentric camera, implemented using the Google Glass platform. The occupant location is established through the implementation of machine-vision techniques to identify reference objects located within the environment that are then cross-referenced against a knowledge base that contains the reference object's known location. The reference objects are identified by fiducial markers placed upon them. Fiducial markers can be defined as artificial landmarks, or reference points, that are added to an environment to aid in tracking, alignment, and identification within the environment [288]. They can either be placed upon a fixed point within the environment to enable a moving camera to allow the location of the camera to be determined or they can be placed on moving objects to allow the location relative to a fixed/moving camera to be determined [289]. Within the context of this work, fiducial markers are defined as images or scenes within the environment that support the alignment, identification, and/or tracking of objects or locations [74]. The need for the occupant's location is a high priority for providing relevant assistance due to the nature of activities being localised to a certain location within the environment; such as making dinner. The presented approach leverages a wearable camera to offer an egocentric view of the environment. This is coupled with fiducial markers placed on "key" objects which offer contextual information as to the occupant's location

In the research presented in this Chapter, the fiducial markers take the form of multiple overlapping shapes applied to "key" objects within the environment, refer to Figure 4.1a for an example of a fiducial marker. These markers were then applied to "key" objects within the environment, as shown in Figure 4.1b. The overlapping shapes were defined through an iterative approach and ad hoc testing to determine the optimum complexity required. A single shape, such as a cross or triangle, did not provide enough unique features points to accurately differentiate the markers from each other. Conversely, overly complex shapes resulted in increased processing time required to identify the shape which resulted in markers being missed due to the speed the occupant would navigate throughout the environment. As feature point algorithms, such as FAST, rely on corner detection methods [290] the shapes were created to maximize the number of corners that could be detected via overlapping shapes which were slightly offset from each other.



(a) Example of a fiducial marker.

(b) An example of a fiducial marker applied to an object.

Figure 4.1: An example of a fiducial marker and how they are applied to objects of interest within the environment.

As shown in Figure 4.1b the markers can be applied to any object, in this case on the telephone. If the telephone is detected, we can determine that the occupant is within the living room and thus can provide the relevant support if/when it is needed within their context. The method of using fiducial markers to identify objects within the environment aims to aid in alleviating some of the traditional problems associated with object detection [291]. One such challenge this method alleviates is attempting to distinguish between multiple identical objects [288], such as kitchen cupboards, as well as negating the requirement to recognise various models of the same appliance that may differ in their appearance, however, offer the same function. Further advantages this method offers is the ability to retrofit it to any object within an environment therefore the need for a fully sensorised environment is no longer required therefore greatly reducing the cost of applying such a system to the occupant's own home. However, there are some negative issues encountered through implementing the proposed solution. Firstly, a unique fiducial marker will need to be generated for each object of interest within the environment. Secondly, effort was needed to ensure that the correct markers are placed on the relevant objects of interest within the environment of interest. The markers can be placed by a non-expert user, however, it would be beneficial if they are placed by a person who is familiar with the occupant's routines in order to ensure all the relevant objects have had a marker attached.

4.2.1 Streaming Data from a Wearable Camera

This research was conducted using the Google Glass device [167], which is equipped with a first-person video camera at eye level, in addition to a full sensor suite such as accelerometer, gyroscope, and GPS amongst others. Chapter 3 provides a fuller overview of Google Glass and its technical specification.

4.2.2 Near Real-Time Streaming

In order to determine the location of the occupant the video feed will be streamed from the Google Glass device showing an egocentric view of the environment. Any fiduciual markers found in the video feed will be detected and the marker's associated location will be logged as the occupant's current location.

In order to offer relevant, timely support the video feed from the Google Glass was processed in real time. This functionality was not supported by Google Glass by default. An app was developed for Google Glass that allowed a video stream to be captured and then sent via Real Time Streaming Protocol (RTSP) to a cloud-based server. The video feed was then freely accessible by multiple sources. In order to process the video, the machine-vision server accesses the video stream via RTMP (Real Time Messaging Protocol) and performs the video processing. This approach did, however, introduce a brief latency (<4 seconds) due to Glass' efforts to lower its temperature during high load situations, such as streaming [9].

4.2.3 Server Offloading

As wearable devices are traditionally "resource poor" in comparison with contemporary server hardware [9] Google Glass was responsible for capturing the video stream and delivery of reminders and notifications only. This was to avoid introducing a large delay within the processing time from detecting a marker within the video stream and determining the location from the marker to establish the location of the occupant. The image processing was offloaded to a server via RTSP for processing, thus decreasing the time taken for object detection and for the appropriate response to be given, along with increasing battery life on the Glass platform. Ha *et al.* carried out a comparison of an assistive application (OCR: Optical Character Recognition). They compared the performance and energy use of Glass performing the recognition task, both via on board processing and comparing this against offloading the processing to a server via a real-time stream from Google Glass [9]; their results are shown in Table 4.1.

Table 4.1: Comparison of offloading *vs.* on-board processing for Google Glass. Mean over five runs, standard deviation shown in parenthesis [9].

Metric	On-Board (seconds)	Offloading (seconds)
Per-Image Energy	12.84 (0.36)	1.14 (0.11)
Per-Image Speed	10.49 (0.23)	1.28 (0.12)

As can be seen from Table 4.1, there is almost an order of magnitude difference in both speed and energy used in offloading compared to on-board processing. Google Glass offers a 2.1V 570mAh (7560 Joule) battery, equating to an 11-minute battery life when performing on-board processing and an 111 minute battery life when offloading to a server, along with an decrease in the processing time required to perform the recognition [9]. Battery life can be further extended with external battery packs, however, with the current rate of advance in battery technology the battery life of future generations of wearable devices will be less of a challenge.

The imagine recognition was carried out using the OpenCV library [292], which is an open source library aimed at real-time macine-vision, using a desktop machine as the server. The technical specification of the server was as follows: Intel Core2Quad (Q9950) 2.83GHz CPU, 8GB RAM. The video was transmitted at 640x480 at 20fps. Due to processing limitations of Google Glass a variable lag (<3s) was introduced on the video stream. This was due to Google Glass's efforts to lower the operating temperature, which is achieved by reducing the clock speed of the CPU. At high temperatures, the Glass firmware limits of the CPU to 600Mhz or 300Mhz to cool down via power reduction [274].

It should be noted that there can be limitations to offloading, in particular, privacy and security. This is due to the data being streamed over a network from the device to a server which can increase the likelihood of unauthorised access or data breaches through malicious attacks. Additionally, the occupant could feel uncomfortable knowing their data is being offloaded to a server outside of their control. Particularly given the sensitive nature of a egocentric video stream within the occupants own home. Certain steps can be taken to reduce the privacy issues, such as in this research were the video feed is not stored, only the I.D. of the detected marker along with a timestamp are stored. Additional considerations that must be taken into account is the potential costs involved with regards to the scalability of an offloading approach should a third party server provider be used. Especially if mulitiple occupants are being supported within the environment as this would require mulitiple video streams to be offloaded and processed.

Within the implementation in this research a desktop server was used within the occupant's environment to reduce issues with data privacy and security. As the data would not be leaving the occupants home network, alongside only recording the detected marker name and timestamp. This approach was also low cost with a low spec desktop PC being required, however, if multiple occupants were being supported a higher specification desktop should be considered.

4.2.4 Comparison of Feature Detection Algorithms

To detect and identify the fiducial markers a feature detection algorithm was required. To determine the best fit for this purpose, a review of the literature pointed towards Orientated FAST and Rotated BRIEF (ORB) as being the best fit for this work [293, 294, 295]. A brief overview and comparison of the algorithms is presented.

SIFT: Scale-Invariant Feature Transform (SIFT) was developed by Lowe [296] as a method of extracting distinctive invariant features from images to provide reliable image matching. The features extracted are invariant to both scale and rotation, in addition to being robust to the effects of affine distortion, noise, and lighting changes [296]. This method also allows for highly distinctive features to be extracted so that a single feature can be correctly matched against a large database of features from multiple images. Image matching is performed by matching individual features against a database of known features using a nearest-neighbour algorithm along with a Hough transform to identify clusters belonging to a single known object. Verification is then performed through least-squares solution for consistent pose parameters [296].

SURF: Speeded Up Robust Features (SURF) was developed by Bay *et al.* [297] as scale and rotation invariant feature point detector and descriptor which relies on integral images for image convolutions. In order to detect interest points an integer approximation of the determinant of a Hessian blob detector is calculated with three integer operations utilising a precomputed integral image [297]. Analysis has shown that it is three times faster than SIFT while performance is comparable to SIFT. The main advantage of SURF is in handling images with blurring and rotation; however, it falls down at handling viewpoint change and illumination change [292].

FAST: Features from Accelerated Segment Test (FAST) was created by Rosten & Drummond [298] to be a low computational method of detecting features in real-time video via the application of machine learning. FAST creates a decision tree which can correctly classify all corners in the training set. To classify a corner a pixel "P" is selected and a circle of 16 pixels is selected around it. Four pixels from the circle are then examined (1 and 9 first, if these are too bright or dark then 5 and 13 are checked) if "P" is a corner then at least three of these pixels should be brighter or darker than "P". While it is several times faster than other existing corner detectors it is not robust to high levels of noise and is highly dependent on a threshold value.

ORB: Orientated FAST and Rotated BRIEF is an alternative to SURF and SIFT which was proposed by Rublee *et al.* [293]. ORB uses FAST (Features from Accelerated Segment Test) in pyramids in order to detect stable key-points and selects the strongest features using FAST. FAST is an efficient method of finding key-points in images. It is a particularly common solution in real-time systems that match visual features, however, it must be augmented with pyramid schemes to take scale into account, and in the case of ORB a Harris corner filter must be added to reject edges [293]. ORB employs the Binary Robust Independent Elementary Features (BRIEF) feature descriptor which employs simple binary tests between pixels in a smoothed image patch and offers robustness towards lighting, blur, and perspective distortion.

ORB implements the intensity centroid method of corner detection as defined by Rosin [299]. ORB features are invariant to rotation and scale, resulting in a very fast recogniser which is robust to viewpoint invariance [294], while being faster than both SIFT and SURF based algorithms while maintaining accuracy [295]. The intensity centroid assumes that a corner's intensity is "o" set from its centre, and that this vector can be used to impute an orientation. A previous study by Gil *et al.* [300] has shown that a strength of ORB is its ability to accommodate low brightness conditions, in part due to ORB implementing the Harris Corner Detection algorithm which Pribyl *et al.* has shown to be robust in low lighting conditions [74].

Tareen and Saleem [10] undertook a study to investigate the computational cost for a range of feature point algorithms by demonstrating the computational cost per feature point based on the mean values from a range of images within multiple datasets as displayed in Table 4.2.

OpenCV [292] provide a hardware and operating system agnostic implementation of various feature point recognition algorithms thus removing hardware

Table 4.2: Computational cost per feature point [10].

Algorithm	Mean Feature Matching Time (μs)
SIFT	142.02
SURF	89.66
ORB	11.82

platform limitations. Being able to run on smartphones, Raspberry Pi, or more powerful desktop devices.

Following review of the aforementioned approaches, ORB was selected as the algorithm of choice. This was due to its lower computational overhead compared with the alternative approaches while still offering a high level of accuracy. Its robustness to lighting conditions also highlighted it suitableness to the application of AAL where the occupant would need to be supported at all times of the day with varying lighting conditions due to differing interior lighting and the natural lighting changing throughout the day.

However, it should be noted that are some limitations when using feature point algorithms for fiducial marker detection. Firstly, there is a reliance on sufficient complexity being present within the marker design in order to reliably differentiate the markers from the background. Additionally, scalability can also be an issue due to the potential limit of the number of unique markers that can be generated without increasing the number of mistaken detections (false positives). Marker placement is also a limitation as factors such as lighting and occlusion can result in reduced performance. Additionally in situations were there may be multiple fiducial markers in close proximity feature point algorithms can struggle to differentiate betwen markers. Occlusion can lead to missed detections or false positives due to the markers being fully or partially occluded within the environment. While this is a limitation of feature point recognition algorithms it can be possible to detect partially occulded markers through the use of deep learning approaches. Noise is also an important factor to consider and can be introduced through lighting, network faults, sensor faults, or artificats introduced through compression. The result of noise within the data can result in reduced performance within feature matching and detection which can increase the rate of false negatives and false positives. Image distortion is another factor that can potentially affect the performance of a feature point detection algorithm. Distortion can be introduced through the camera lens affecting the detection of the fiducial markers, Chapter 5 discusses how the camera was calibrated to take account of this distortion.

4.2.5 K-Nearest Neighbour Matching

A KNN algorithm [301] was used to match the detected feature points of a marker against the known marker templates to determine if a marker is present. A KNN algorithm can be formally defined as finding the K closest (similar) features to a query feature among N points in D-dimensional feature space [302]. In the presented implementation, a simple, from a reasoning perspective, version of a KNN is used, a Brute Force Matcher which takes a descriptor of one feature in the first set which is then matched with all the other features in the second set using a distance calculation with the closest match being returned.

Cheng *et al.* bench-marked multiple algorithms for the purposes of image matching which shows how the Brute-Force matcher compares to other feature matching algorithms [303]. A Brute-Force Matcher may be one of the worst performing matchers [303] in terms of time taken to establish a match, though the detection time as implemented in this research is less than one second, it was concluded that it is the best performer in terms of accurately identifying the correct matches [303].

The Brute-Force Matcher has been used in this research to compare feature points for matching pairs. For each feature in the object, the Brute-Force Matcher locates the closest feature between two pairs by trying every one. The similarity between two pairs is represented by the Norm Hamming distance. This was more efficient, in terms of computation speed, than alternatives as Norm Hamming distance can be implemented using an XOR followed by a bit count which can be carried out extremely fast on modern CPUs [304]. A minimum Hamming distance is set to ensure that only good matches are selected. A match is considered good when the distance is less than three times the minimum Hamming distance set. An overview of the process of setting the minimum and maximum distance along with the good match selection pseudo-code is presented in Algorithm 1.

Algorithm 1 The process of setting the minimum and maximum distance along with the process of selecting a good match.

$mindist = 100$
$maxdist = 0$
$dist = \emptyset$
$matches[\emptyset]$
for $matches$ **do**
  if $dist < min\_dist$ **then**
    $min\_dist = dist$
  end if
  if $dist > max\_dist$ **then**
    $max\_dist = dist$
  end if
end for
for $matches$ **do**
  if $3xmin\_dist < matches : distance$ **then**
    $goodMatches[matches]$
  end if
end for

However, it should be noted there are some limitations with utilising a KNN algorithm. One such limitation is that of computational complexity due to the need to calculate the distances between data points, this limitation increases in cost as the dataset increases in size. This also limits the use of KNN to hardware with sufficient memory storage for storing the entire dataset and thus it is limited due to memory consumption in relation to the size of the dataset. KNN can also be sensitive to noise within the dataset, for example, if there are any outliers within the dataset this can have an effect on the result of the nearest neighbour calculations due to outliers potentially being treated as neighbours which may distort the boundary. Noise within the data can also result in the noise being mistaken for a neighbour resulting in misclassification. Additionally, the choice of the value of K can have a significant impact on the results. A low value of K can result in the model being adversly affected by individual data points, particularly if the dataset contains outliers, which can result in the model being overfitted to the dataset. A high value of K can result in the model becoming too general and can potentially lead to the model underfitting the data. This is due to the high K value casuing the algorithm to consider a larger number of neighbours and can result in reduced accuracy.

4.2.6 Two Stage Filter

During initial testing a high number of False Positives (where a tagged object is determined to be present when it is not) were occurring. These were found to be caused by objects in the environment containing a partial match to the fiducial markers. In order to dismiss the number of FP reported by the system a two-stage filter was used. For the first stage, the homography was used as a model for correct matches allowing a transformation to map the points in the template image to the corresponding points in the frame. The number of inliers that contributed to the homography were determined and compared against a threshold value (refer to Algorithm 1). If the number of inliers matched or exceeded this value, then it is passed onto the second stage.

The second stage employed a Vote Function where any further FP that have passed through the first stage are removed. A batch of frames (three in this implementation) were processed. The object most likely to be present in each frame was determined and stored. Once the most likely object for each frame has been determined a vote count is performed. Once this count passed a pre-determined (defined by a human expert) threshold value the most likely object was determined to be present. The pseudo-code for the second stage filter is presented in Algorithm 2. Figure 6.1 illustrates how these multiple algorithms were combined as a whole system.

Algorithm 2 Vote function combining multiple frames to determine if an object is detected.

$threshold = \epsilon$
$objectID[\emptyset]$
for $totalNumberOfObjects$ **do**
  if $detectedObject == objectID$ **then**
    $objectID[detectedObjectCount++]$
  end if
  if $objectID[detectedObjectCount] == threshold$ **then**
    $objectDetected$
    $objectID[\emptyset]$
  end if
end for
return $ObjectDetected$

It should be noted that there can be potential limitations with the threshold value being set by a human expert. Firstly, a potential limitation is that is that of subjectivity as different human experts could have differing opinions on what is an appropriate threshold value. Additionally, a threshold value being set by a human expert may cause a lack of generalistaion, due to the threshold value being set based off a familiar dataset which may not generalise to further datasets. The threshold being set by a human expert can also result in a bias being introduced to the system due to the human expert's bias towards a certain instance or class within the data. Lastly, it can be a time consuming process to manually set a threshold value which may not be scalable as the dataset increases in size along with reducing the level of transparancy within the decision making process.

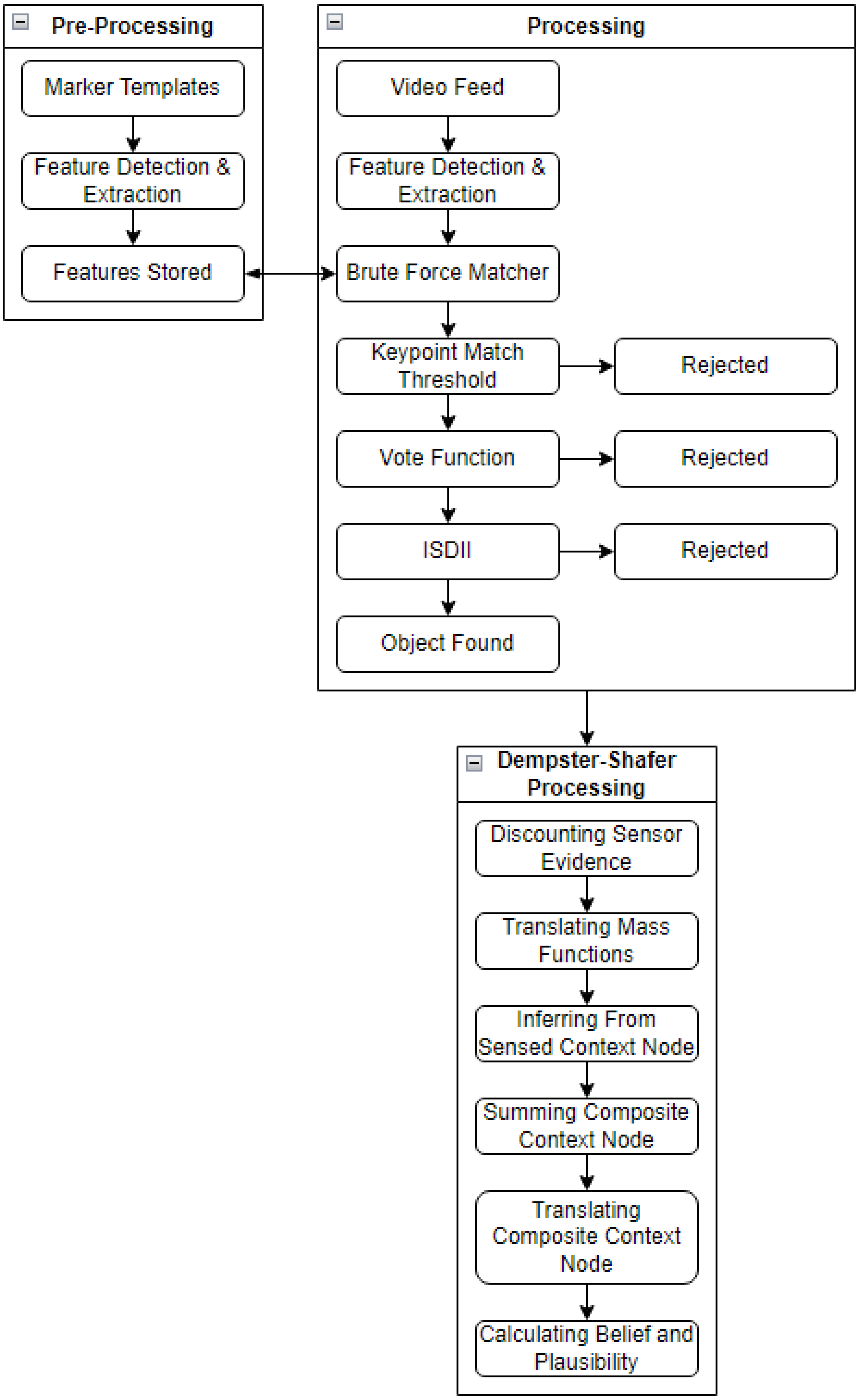


Figure 4.2: Overview of the flow of data, showing the two stage filtering process.

4.2.7 Benchmarking

To assess the performance of the proposed work in this thesis a comparative technique was required as a benchmark. To address this requirement a dense sensor-based solution was used to provide a comparison with the machine-vision system. The dense sensor system consisted of TyneTec binary contact sensors placed on the 'key' objects that also had a fiducial marker attached to them. There was a total of 14 TyneTec sensors, events were uploaded to a MySQL database for retrieval. Further details of the objects the TyneTec sensors were attached to and their location within the PCRC and UJAmI labs can be found in Chapter 3.

The data for both the dense sensor and machine-vision datasets were collected by a single researcher simultainously. The TyneTec sensors and fiducial markers were placed within the environment while the researcher performed the study protocol while streaming from the Google Glass device. The machine-vision video stream was then process in near real-time to detect any fiducial markers within the stream and then stored the marker I.D. and timestamp. This allowed a comparision to be made between the accuracy of the dense sensor based system in comparision to the machine-vision based system by comparing the number of false negative and false positive events.

4.2.8 Activities of Daily Living Study Protocol

As was introduced in Chapter Three, a range of nine unique activities were repeated within three differing routines (activities were duplicated in both PCRC and UJAmI) that were representative of daily routines [66], with the goal of recognising the component locations that make up each activity. If prepare/drink water is taken as an example activity, then the component locations would be the kitchen door, the cup cupboard, the tap, and then finally the kitchen door again. Three routines, specified in Chapter 3, were carried out. The first containing ten activities and the remaining two containing eleven activities (routine two and three contained repeating activities). The first routine did not contain the phone call activity to simulate phone calls being a typically unscheduled activity in the real world. These ranged from simple activities such as drinking a glass of water to more complex activities, such as preparing hot food. The activities are presented in Table 4.3, with the full routines presented in Table 4.4.

These routines were performed under the same lighting conditions (brightly lit with artificial lighting and partially closed window blinds) to minimise any potential discrepancy between identical activities in differing routines. To promote the accuracy of the machine-vision and binary sensor location systems, the ground truth was obtained from a time stamped video. The occupant's location reported from the location systems were then compared to the ground truth from the video. Both the vision and TyneTec data were gathered at the same time, as the researcher carried out the activities while wearing Google Glass. The objects were also fitted with TyneTec sensors, allowing the data to be gathered within the one routine. All components of the system were time synced with the MySQL server to ensure that the events were synchronised. Lighting was controlled as discussed in Section 3.3.4.

Table 4.3: Full list of activities that were performed during the three routines, these were chosen to represent a range of ADL that take place within a kitchen/living room area.

Activity Number	Activity
1	Prepare/drink water
2	Prepare/drink tea
3	Prepare/drink hot chocolate
4	Prepare/drink milk
5	Make/receive phone call
6	Prepare/eat cold meal
7	Prepare/eat hot meal
8	Watch TV
9	Wash dishes

In order to assess the viability in applying the proposed solution to multiple environments the aforementioned routines were carried out in a second location, the UJAmI smart lab, University of Jaén. Ceiling lighting and window blinds were used to control the lighting conditions. Additionally, activities remained the same within each routine along with both the markers and wearable sensor, the only variable being the environmental layout. Ground truth was gathered by the researcher involved from manually annotated video data to ensure the accuracy of the vision system. As it was the viability of the vision system that was of interest only the vision results were compared between the results of the experiment in PCRC and UJAmI.

The resulting true positives, false negatives, and false positives from both the dense sensor and machine-vision systems were compared as a means to evaluate the performance of the two systems. Additionally, the recall, precision, and F-measure for both systems were calculated to provide an additional means to evaluate and

Table 4.4: Breakdown of activities that took place in each routine.

Routine 1 (R1)	Routine 2 (R2)	Routine 3 (R3)
3	4	3
1	6	1
7	1	5
9	5	7
8	1	1
1	2	8
8	8	2
6	7	8
9	9	6
1	8	9
N/A	1	4

compare the performance of both systems.

4.3 Results

This Section describes the results of the machine-vision localisation system, along with details of the results from the dense sensor system when compared with the ground truth from the annotated video data. Due to the high number of true negatives (TN) over twenty thousand, from the machine-vision system a skewed dataset was produced. Due to this the performance was assessed by measuring recall, precision, and F-Measure. These were focused on to avoid misinterpreting the high number of TN giving an incorrect weighting to the results.

The results from the machine vision system at the PCRC lab are presented in Tables 4.5 and 4.6, and the results from the UJAmI lab are presented in Tables 4.7 and 4.8. Tables 4.6 and 4.7 show a total of nine FP from the 350 total events these were due to a mistake being made in recognising the fiducial markers and detecting them as a different marker.

As shown in Table 4.9 there was a total of 32 FN (175 total events) within the PCRC lab. The majority of these (16) were due to corruption within the video frame during transmission. The remaining FNs where due to varying reasons, such as missing frames. There were a total of 59 FNs within the UJAmI lab (Table 4.10), most of these (47) were due to the camera auto-focus failing to focus. This could be seen as a weakness of the system as if the camera did not have enough time to focus on the marker then they may be missed. The rest of the FNs were due to varying reasons, such as missing frames due to network latency or corrupted frames.

Table 4.5: Results of Recall, Precision, and F-Measure for the machine vision-based system – PCRC.

Routine	Total Events	Recall	Precision	F-Measure
R1	58	0.74	0.98	0.84
R2	56	0.88	0.94	0.91
R3	61	0.84	0.96	0.89
Total	**175**	**0.82**	**0.96**	**0.88**

Table 4.6: Breakdown of machine vision sensor classification outcomes including TP, FN, and FP – PCRC.

Routine	Total Events	#TP	#FN	#FP
R1	58	43	15	1
R2	56	49	7	3
R3	61	51	10	2
Total	175	143	32	6

Table 4.8 and 4.10 presents the machine-vision results from the UJAmI lab. As shown in Tables 4.8 and 4.5 there is reduction of the average Recall and F-Measure by 0.16 and 0.09 respectively with a rise in Precision of 0.01, suggesting that it is viable to apply the system to multiple environments. Even though there was a drop in performance in terms of F-Measure and Recall the system was still able to accurately determine the occupant's location. The results from the binary contact sensors are presented in Tables 4.11 and 4.12. While the binary contact sensors provided more accurate results this does not fully demonstrate

Table 4.7: Breakdown of machine vision sensor classification outcomes including TP, FN, and FP – UJAmI.

Routine	Total Events	#TP	#FN	#FP
R1	58	39	19	1
R2	56	38	18	1
R3	61	39	22	1
Total	175	116	59	3

Table 4.8: Results of Recall, Precision, and F-Measure for the machine vision based system – UJAmI.

Routine	Total Events	Recall	Precision	F-Measure
R1	58	0.67	0.98	0.80
R2	56	0.68	0.97	0.80
R3	61	0.64	0.98	0.77
Total	175	0.66	0.97	0.79

the additional advantages the machine vision system provides over dense sensor placement.

One of the key advantages that the vision methods offers which was uncovered during the experiments is that interaction with an object is not required to determine the occupant's location within the environment. This can offer a timelier location update compared to dense sensor placement. In the experiments, the occupant's location was reported before they had interacted with the object thus offering a timelier update. This was due to the manner in which each system reported an event, with the dense sensor placement an event can only be reported as the occupant is interacting with the object of interest. With the vision-based system the interaction could be reported before the occupant has physically interacted with the object, being able to recognise the intention of interaction as the occupant approached the object. Also, if the occupant became confused or decided not to use the object their location would still be captured. This would have otherwise been lost in a traditional sensor based smart environment. Another potential advantage is that of multiple occupancy. As each occupant will use a

Table 4.9: A breakdown of FN machine vision events – PCRC.

Cause	#FN
Corrupt Frame	16
Other	8
Unknown	8
Total	32

Table 4.10: A breakdown of FN machine vision events – UJAmI.

Cause	#FN
Unfocused	47
Unknown	12
Total	59

wearable device it would be possible to locate each occupant within the environment and to infer their activity from their own first-person view. Nevertheless, this is working under the assumption that only the occupants of the environment will require support, as any visitors will not have a wearable device. If any sensor activity is detected without a corresponding machine-vision event, then it would be assumed that the visitors have activated a sensor and thus that event should be ignored. While it is possible for the machine-vision system to miss an event, there would be opportunities for this event to be detected due to the constant monitoring of the environment through a camera. As the vision system does not require interaction even if the initial event is missed, follow up events may still be captured. This additional information is lost in a traditional dense sensor environment and once the occupant has finished interacting with the object there is no longer any opportunities to detect a follow up event.

4.3.1 Application to Multiple Environments

This study also investigated the viability of translating this solution to other environments. Occupants generally should be supported within their own home which needs to be taken into consideration when developing a solution to that of AAL. The proposed system offers reduced financial costs in terms of initial equip-

Table 4.11: Results of Recall, Precision, and F-Measure for the dense sensor based system.

Routine	Total Events	Recall	Precision	F-Measure
R1	58	1.00	1.00	1.00
R2	56	0.93	1.00	0.96
R3	61	0.90	1.00	0.95
Total	175	0.94	1.00	0.97

Table 4.12: Breakdown of dense sensor classification outcomes including TP, FN, and FP.

Routine	Total Events	#TP	#FN	#FP
R1	58	58	0	0
R2	56	52	4	0
R3	61	55	6	0
Total	175	165	10	0

ment purchase and maintenance, along with a reduction in the invasiveness for the installation compared to traditional indoor localisation methods as discussed in Chapter Two. Details on the costs of purchasing the relevant equipment and installation can be found in Table 4.13 *vs.* the Google Glass Explorer edition cost of approximatly £1,200 at the time of writing. The issue of multiple occupancy is also addressed as this solution allows individual support to be given to each occupant as they have a unique first-person view of the environment. This does, however, assume that only the occupants require support and that any visitors to the environment can be assumed to not require any assistance allowing support to be given in the form of notifications/reminders to assist with completion of ADL. This solution aims to improve context aware support through the localisation of objects within a smart environment.

One aspect of AAL that must be taken into consideration is the acquisition and maintenance costs of implementing a sensorised environment. A large network of embedded sensors is normally required which results in a system that is costly to maintain, relatively obtrusive (as sensors are required on every interactable object),

Table 4.13: A breakdown of approximate costs with associated sensor platforms [11].

System	Cost	Installation
Elk M1	£5,000	DIY
Lagotek	£5,000	DIY
Control4	£50,000	DIY
Control4	£98,000	Professional
X10	£250	DIY
Creston	£49,000	Professional
EIB Instabus	£11,000	Professional
KNX	£25,000	Professional

and sensitive to the performance of the sensors [98]. Table 4.13 presents the estimated costs involved in implementing both dense sensor and fixed video camera systems within a household. As can been observed from the Table 4.13 there is a high financial cost involved in the purchase and installation of traditional methods of indoor localisation. While a DIY installation goes a long way to reduce these costs (Control4 price is reduced by approximately £57,000 from the professional installation), it must be considered that the occupants that would benefit from such as system may not be physically or mentally fit to carry out such an intensive installation. An additional advantage towards the proposed system, and vision systems in general, is that generic hardware can be used for multiple applications to aid of AAL [98].

4.3.2 Multiple Environments

The results from the experiment in the UJAmI lab offer an insight into the viability of applying the system to other environments. The results support the hypothesis that a single wearable camera allows occupant tracking within an environment with the goal of determining location, subsequently showing consistent results across multiple environments. As the markers are placed on common objects that are ubiquitous to every home environment, the markers used in the PCRC experiment could be directly used when recreating the experiment in the UJAmI lab without modification. This facilitated a simple and fast set up time (five minutes) compared to traditional methods such as dense sensor placement or the installation of static cameras [188, 191]. Due to the small nature of the dataset, missed events have a larger impact, resulting in a reduction in recall and F-Measure, however, the precision was increased which is significant as it is important that the occupant's location is correctly identified to offer relevant support. Despite this the results suggest that the method is viable across multiple environments. The creation of a larger dataset is warranted to gain a more accurate picture of the performance.

4.4 Discussion

The contributions offered by this Chapter include addressing a problem previously identified with that of wearable devices such as Google Glass. That is, that their impact in ubiquitous computing and ambient intelligence systems has been partly slowed by their lack of streaming [187]. This has been addressed in Section 4.2.1 by the development of live streaming functionality from a wearable device, Google Glass in this case, which allows the video stream to be accessed by multiple sources using a media server.

Near real-time vision based indoor localisation through an egocentric camera utilising fiducial markers. This alleviates the issues identified within Chapter Two, such as occlusion from fixed cameras where the occupant is not within the camera's field of view due to large objects occluding the occupant or "blank" areas of the environment where the camera's field of view does not cover. While there is a risk of occlusion of the fiducial markers this is greatly reduced through the use of a first-person camera which removes the issue of covering the entire room along with large items, such as doors/fridges, occluding the object of interest.

An additional advantage the system offers is avoiding the need to be trained to each environment that it is to be deployed within by using fiducial markers. This allows the system to be quickly and easily deployed within new environments in comparison to implementing traditional methods of indoor localisation.

Due to the system operating in near real-time it does not encounter the same issues as intermittent image capture system. Where vital information could be lost if the occupant interacts with an object or navigates throughout the environment. In the previously discussed works the method of image capture relied on intermittent captures, *e.g.* at set time intervals 30 frames were captured. This could cause vital information to be lost as object interactions may have taken place within the time period were the system was not capturing information. As the presented system operates in near real-time every frame is being processed, therefore vital information will not be lost through intermittent image capture.

The proposed approach offers other secondary advantages when compared to a traditional method of indoor localisation, such as dense sensor placement, that are unique to this method. Such as the first person view and lack of required interaction and multiple occupancy, where each occupant that requires support need only to wear a device to obtain their unique first person viewpoint and the information on what objects they are interacting with. Additionally privacy is preserved as the video stream is not viewed by anyone with the detection events being the only information which is stored on the server.

4.5 Conclusion

A method of indoor localisation is presented utilising a wearable camera to determine location based upon objects viewed within a scene. This was compared with a traditional method of indoor localisation (dense sensor placement) employing annotated video data as the ground truth. Thus, it supported the hypothesis that the use of a single wearable camera allows occupant tracking within an environment with the goal of determining location. While the machine-vision results were found to be less accurate than dense sensor placement, they demonstrated that the proposed method is viable and offers other secondary advantages that are unique to this method, such as the first-person view and lack of required interaction.

Further, the work presented demonstrated the viability of applying the solution to differing environments. The performance of the system at the UJAmI lab were comparable with the previous experiment carried out at the PCRC lab. With the UJAmI experiment showing an average recall, precision, and F-measure of 0.66, 0.97, and 0.79, respectively in comparison to the PCRC experiment results of recall, precision, and F-measure of 0.82, 0.96, and 0.88, respectively. The duplication of the experiment in UJAmI demonstrated the viability of applying the solution to multiple environments which has been shown to be a challenge within the domain of AAL, as was discussed in Chapter Two. The lack of training, use of common objects and hardware are attributed to this success. Additional advantages of this approach is the ability to generalise to other users due to the lack of personalisation required. This is due to the system requiring markers be placed on key objects with no input being required from the user once the application is started. This is of particular importance when it comes to older users who typically have lower levels of confidence with regards to the use of technology.

There were, however, some limitations of using such as static approach to storing the object's location within a knowledge base, such as objects being moved or certain objects that may not have a static location, for example personal devices. As the object's location is assumed to be fixed within the environment and used as "key" objects within each room. If the location of any of these objects is changed without being updated within the database, the accuracy of the system will suffer. This is further compounded with "personal" devices, such as a smart phone, as they do not have fixed location and therefore cannot be relied upon to find a location update for the occupant. However, these personal devices can still be leveraged to gain an understanding as to the activity that the occupant may be carrying out. Another limitation inherent with wearable camera solutions is that they rely on an "always-wear" approach as the system is reliant on the occupant to remember to put the Glass on in the morning. This is somewhat mitigated in that 74% of the adult population wear corrective lenses [305] and with the ability to insert prescription lenses into Google Glass. It could replace their normal glasses to try and avail of their daily routine of wearing glasses. Additional limitations include the potential for false positives within the environment, these can be caused by complex scences where there may be a number of fiducial markers within the video stream. False positives can also be caused when the occupant is navigating throughout the environment as fiducial markers could remain within the FoV even when the occupant is not at that location. A further limitation is the potential for false negatives which can be caused by corruption within the data stream or via external factors such as lighting or occulsion. Chapter Six will involve determining activity based on the objects located within the field of view, along with mitigiating another limitation of the system were false positives could be generated from the occupant navigating through the environment or through general gaze activity.

4.6 Associated Publications

Shewell, C, Nugent, C, Donnelly, M, Wang, H & Espinilla, M 2017, "Indoor Localisation Through Object Detection Within Multiple Environments Utilising a Single Wearable Camera", Health and Technology, vol. 7, no. 1, pp. 51-60. https://doi.org/10.1007/s12553-016-0159-x

Shewell, C, Nugent, C, Donnelly, M & Wang, H 2016, "Indoor Localisation Through Object Detection on Real-Time Video Implementing a Single Wearable Camera". in International Federation of Medical and Biological Engineering. vol. 57, Springer, pp. 1231-1236, 14th Mediterranean Conference on Medical and Biologi-

cal Engineering and Computing, MEDICON 2016, 17/09/16. https://doi.org/10.1007/978-3-319-32703-7_237

Chapter 5: Optimising Marker and Object Detection Through Enhanced Filtering and Segmentation

5.1 Introduction

Chapter 4 assessed the technical feasibility of leveraging a wearable camera to provide an egocentric view of the immediate environment, coupled with fiducial markers placed on "key" objects to allow the approximation of objects and occupant location within an environment. This chapter refines this approach by assessing the ORB algorithm against another fiducial marker detection algorithm, ArUco. As well as generating a method to filter out additional FPs that are caused by the occupant's navigation of the environment or through general gaze activity. During a collaboration with the University of Jaén the ArUco algorithm [197] was proposed as a potential improvement over the ORB algorithm, presented in Chapter 4. ArUco was chosen as a comparision algorithm for a number of reasons. Firstly, it has been developed as a dedicated fiducial marker detection algorithm and is an open-source and widely adapted within the computer vision community. ArUco also supports a wide range of programming languages along with an accssible API for the creation and detection of fiducial markers. ArUco is also optimised for real-time marker detection which is key given the requirements of supporting an occupant at home with their ADLs. In order to assess the algorithms, recordings were captured of an occupant carrying out a set of ADL, using Google Glass, introducing levels of varying motion blur and lighting conditions. To promote a fair comparison, the two-stage filter system, described in Chapter 4, was not applied and instead all video was processed on a frame by frame basis.

One challenge that was observed during testing of the system in Chapter 4 was the detection of FPs when an occupant was navigating throughout an environment or FPs arising from the wearers general gaze activity as they interacted with objects during the completion of activities. To address this issue, ISDII was developed. This chapter details the rationale, methodology, testing, and evaluation of both the ORB and ArUco algorithms alongside the ISDII system.

5.2 Methodology

This Section details the methodology adopted to develop the system. The design of the fiducial markers that were used to identify the objects are presented along with a detailed overview of the algorithms used in the evaluation of the system. The system identifies “key” objects within an environment that allows the location of the occupant to be inferred. For example, the detection of a kettle can allow it to be inferred that the occupant is in the kitchen. The detection of “key” objects within the video stream can also allow the current activity to be determined, with the end goal of offering support to occupants’ carrying out their ADLs. Via assisting the occupant in carrying out the activity or alerting carers to abnormal activity levels. A description of the feature point identification method along with the implemented matching process is also presented.

The initial approach compared the performance of two “off-the-shelf” algorithms for performing fiducial marker recognition. Figure 5.1 illustrates the general sequence of events and presents: (i) frames returned from the wearable vision sensor; (ii) fiducial markers located within the returned frames; (iii) the degree of occupant-object interaction as a quantifiable metric.

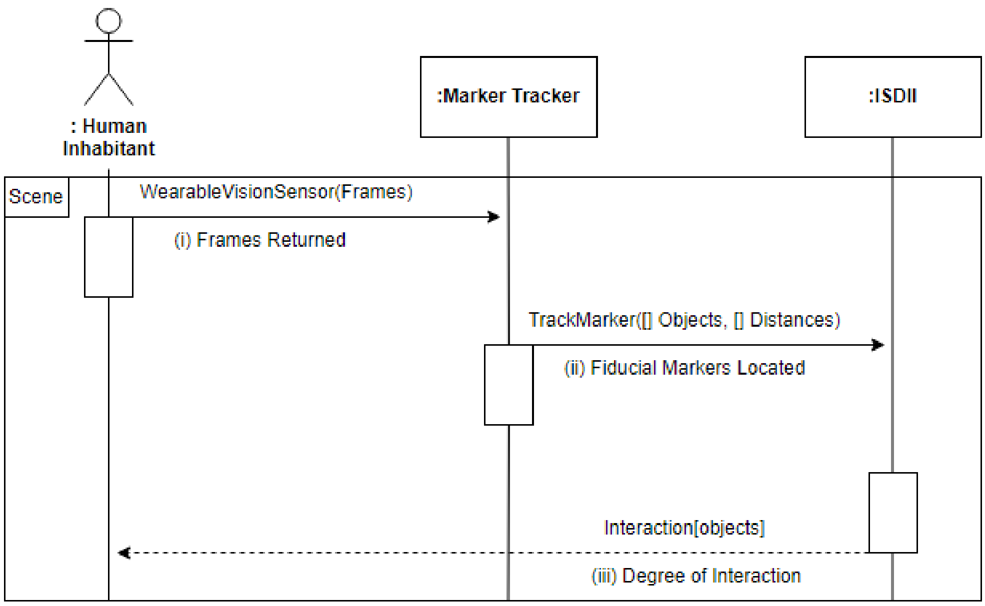


Figure 5.1: Sequence diagram of the wearable vision sensors in ADLs.

Google Glass (Explorer) was employed to provide a first-person view of the occupant's environment. Google Glass facilitates the recording of high definition video (1280x720) and accepts audio based commands from wearers of the device via natural spoken language commands. Pertinent information can also be presented to the wearer via a small prism display that is located directly on the glass in front of the eye.

Traditionally, the uptake of wearable computing devices has been partly slowed by their lack of streaming [187]. In an effort to overcome this, a "Glass App" was developed in our previous work [306], as presented in Chapter 4. A Glass app that supports transmission of live video to a cloud-based server via RTSP. This approach does, however, introduce a short latency between (<4 seconds) due to Google Glass in-built mechanism to lower its hardware temperature during high load situations, such as live-streaming. This results in a reduction of the clock speed of the CPU, thereby reducing the processing rate [274].

Each fiducial marker has a custom identifier applied to it to represent the object it is associated with. The markers were installed on objects of interest throughout the environment with the marker positioned so it fell within the occupant's FoV when the object was interacted wtih. The objects of interest were situated in a location to better represent a real living environment, whilst this resulted in scenes were multiple fiducal markers were present in the video stream Chapter 6 discusses how this challenge was dealt with. The occupant's location is then estimated by means of a 3D reconstruction method that incorporates the known size of the markers, along with the calibration parameters of the vision sensor. Occupant location is of key importance when supporting ADL; in the presented work distance is estimated to determine the degree of occupant-object interaction. Two feature point algorithms were employed to detect the markers located in the environment, using inputs from the vision sensor. What follows is a brief description of the algorithm's main features.

The first method employed the OpenCV implementation of the ORB algorithm for both feature detection and description. This method was developed by Rublee *et al.* [293], and implements FAST in pyramids to facilitate the detection and selection of stable key-points. ORB implements the intensity centroid method of corner detection as defined by Rosin [299].

A Brute Force algorithm (K-Nearest Neighbour) [303] was implemented as a feature point matcher to determine if a marker is present in the frame. A formal representation of a K-Nearest Neighbour algorithm locates the K nearest features to a query feature N points in a D-dimensional space. Even though a Brute

Force matcher is often found to be one of the worst performing algorithms, in terms of time taken to resolve a match, it often provides high levels of accuracy in identifying the correct matches. This finding was reported by Cheng *et al.* [303], which benchmarked multiple techniques for the purposes of image matching. Within this implementation for each feature in the marker, the matcher locates the closest feature in the scene by systematically trying each feature point. The similarity between feature points is represented by Norm Hamming distance. A minimum distance was set to ensure good matches are selected: a match is deemed to be good when the distance is less than three times the minimum distance set.

In order to reduce the number of FP found by the algorithm, a key-point match threshold was used, where the number of inliers that contributed to the homography was calculated and compared against a threshold value [264]. If the number of inliers met or exceeded the threshold then a marker was deemed to be present. A strength of the approach is that the markers can be freely designed. Figure 5.2 offers an example of a custom made ORB marker created by overlapping geometric shapes alongside a pre-made ArUco marker.

The ArUco algorithm is developed under Open Source license: the Berkeley Software Distribution. It has been deployed in several research and enterprise projects[1,2]. ArcUo was developed around the concept of fiducial markers [197]. The markers are automatically generated by ArUco by means of a marker dictionary [307] and focus on extracting the binary code from the rectangles that make up the fiducial marker as presented in Figure 5.2. The processing involves image segmentation, based on local adaptive thresholding. In order to increase robustness to varied lighting conditions contour extraction and filtering is applied, the marker code is then extracted to obtain the internal binary code, and dictionary based correction applied once the binary code is extracted.

[1]http://www.vision4uav.com/?q=node/386

[2]http://vision4uav.eu/?q=researchline/seeAndAvoid_CE_MFandRules

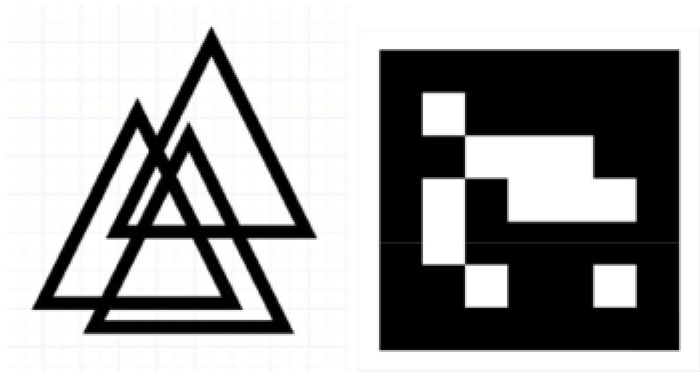


Figure 5.2: A) Example of ORB fiducial marker. B) Example of ArUco fiducial marker.

5.2.1 Intelligent System for Detecting Inhabitant-Objects Interactions

During testing of the vision algorithms, as described in Chapter 4, it was discovered that FP were being generated through general gaze activity. This was due to the occupant looking around the environment when locating an object of interest. Further FP were generated through the occupant's navigation of the environment as various objects came into their field of view as they moved through the environment. An intelligent filter was developed with the aim to detect the degree of interaction between the occupant and the object, based around the observation that when the occupant is interacting with an object of interest they are assumed to be in a close proximity with that object. This also aids in taking account of the differing forms of interaction that certain objects require, namely passive or active interaction. Those objects that require active interaction, such as a microwave, will have a much closer distance threshold compared to those passive objects which are interacted with from a larger distance; such as viewing TV. The filter is known as the ISDII. ISDII uses a two-stage filter in order to manage the uncertainty introduced through FP detections. It is able to determine if an occupant-object interaction is a TP or if it was generated through navigation/gaze activity in real time. It also takes into account the differing forms of interaction that objects may have, for example making a phone call is an active interaction as the occupant has to be in very close proximity to the phone in order to dial the phone number. This is opposed to watching TV which is a passive activity as the occupant would be viewing the TV from a much larger distance than takes place with normal activity object interactions. The output from the marker detection algorithms serve as the input for the ISDII system. These consist of a unique ID associated with the detected markers and the distance of the occupant to the marker. A three stage process is employed:

1. The first stage is to collect and analyse the scenes where interaction occured between the occupant and the object.

2. Thresholds are then determined by a technical expert, establishing the distance at which occupant-object interaction is known to be occurring.

3. Once the threshold distances have been established, ISDII is able to identify interaction on a real-time basis.

In order for ISDII to recognise if occupant-object interactions are occurring, a preliminary threshold value was estabilshed by a technical expert. An initial process was carried out that consisted of recording scenes where an occupant interacted with a series of objects throughout the environment and threshold distances were then set by a human expert, a sequence diagram detailing this step is presented in Figure 5.3. This allows ISDII to calculate, in real time, the distance between the occupant and the object and determine whether an interaction is taking place; the pseudo-code is presented in Algorithm 3.

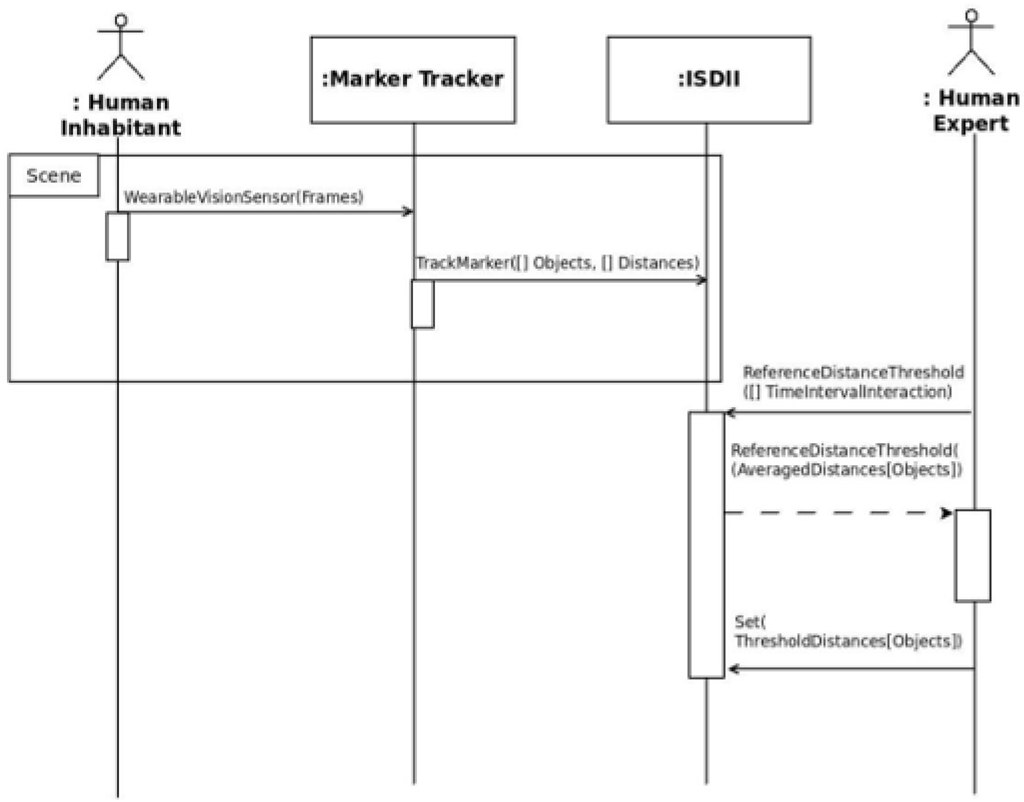


Figure 5.3: Sequence diagram of studying scenes of user-object interactions.

When estimating object interaction in real time scenes, uncertainty is introduced due to missed marker detections in the video stream and measurement errors introduced by the algorithms. In order to manage this uncertainty a two stage filter was developed. The first stage was to remove the high frequency noise using a low-pass filter.

Algorithm 3 Estimation of reference distance threshold to objects.

```
distances = ∅
detections = ∅
for marker ϵ detectedMarkers do
    for interval ϵ interactionIntervals do
        if marker.time ϵ interval then
            distances[marker.object]+ = marker.distance
            detections[marker.object] + +
        end if
    end for
end for
threshold = ∅
for object ϵ objects do
    threshold[object] = distances[marker.object]/detection[marker.object]
end for
return threshold
```

The exponential smoothing [308, 309], is defined in equation 5.1:

$$s_0 = d_0, s_t = \omega_0 d_t + (1 - \omega_0)s_t - 1, \omega_0 \epsilon [0, 1] \tag{5.1}$$

Where d_0 is the initial distance to a marker, t is the temporal index $\epsilon[0, N]$ being N the final size of the set of distances, st is the filtered output, dt the measured data, the distance from the marker, and $\omega 0$ is the smoothing factor (initially set to 0.2); this method has been widely used in control applications [310, 311].

The second filter was designed to mitigate two main causes of FP, removing isolated detections where a marker is detected due to general gaze activity. This is done in order to fit the window of interaction to the true occupant-object interaction, *i.e.* removing the preceding time where the occupant is approaching the object and the proceeding time where the occupant is finished interacting with the object. In order to achieve this, a fuzzy membership function was developed. Fuzzy logic [312] has previously been successfully applied in sensor based signal processing applications [313]. In the context of fuzzy logic the semantics of the linguistic terms are given by fuzzy sets; where the membership degree of the elements x of the base set X in the fuzzy set $A, \mu_{\tilde{A}} : X \to [0,1]$ is defined. The smoothing distance of the markers from the first stage was evaluated by the fuzzy membership function which describes the linguistic term "there is interaction with".

For each object, o_i, a membership function $\mu_{\tilde{O}_i}$ is defined which evaluates the distance between the occupant and the object s_t into a degree of occupant-object interaction between $[0,1]$. The membership function is parameterised by the threshold value of the object d_{oi}, and two weighted factors, ω_1 and ω_2, representing the lower and upper cut-off threshold for interaction respectively, (as presented in Figure 5.4).

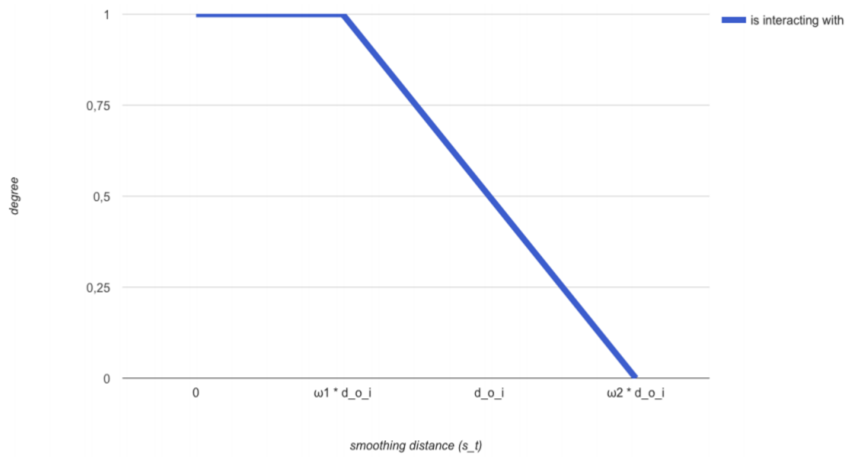


Figure 5.4: Membership function to obtain the degree of interaction with an object.

ISDII provided a degree of interaction representing the occupant-object action within the environment. It should be noted that an upper threshold can be applied using $\alpha - cut$ between $[0,1]$ above which an interaction is determined to have taken place. Pseudo-code for the second stage filter is presented in Algorithm 4 along with a sequence diagram presented in Figure 5.5.

Algorithm 4 Detecting Object Interaction.

$degree = \emptyset$

$detection = \emptyset$

for $marker \epsilon detectedMarkers$ **do**

$distance[marker.object] = \omega_0 \cdot marker.distance + (1 - \omega_0) \cdot distance[marker.object]$

$degree[marker.object] = \mu_{\tilde{O}_i}(distance[marker.object], threshold[marker.object])$

end for

for $object \ \epsilon \ objects$ **do**

if $degree[object] < \alpha$ **then**

$detection[object] = true$

end if

end for

return $[degree, detection]$

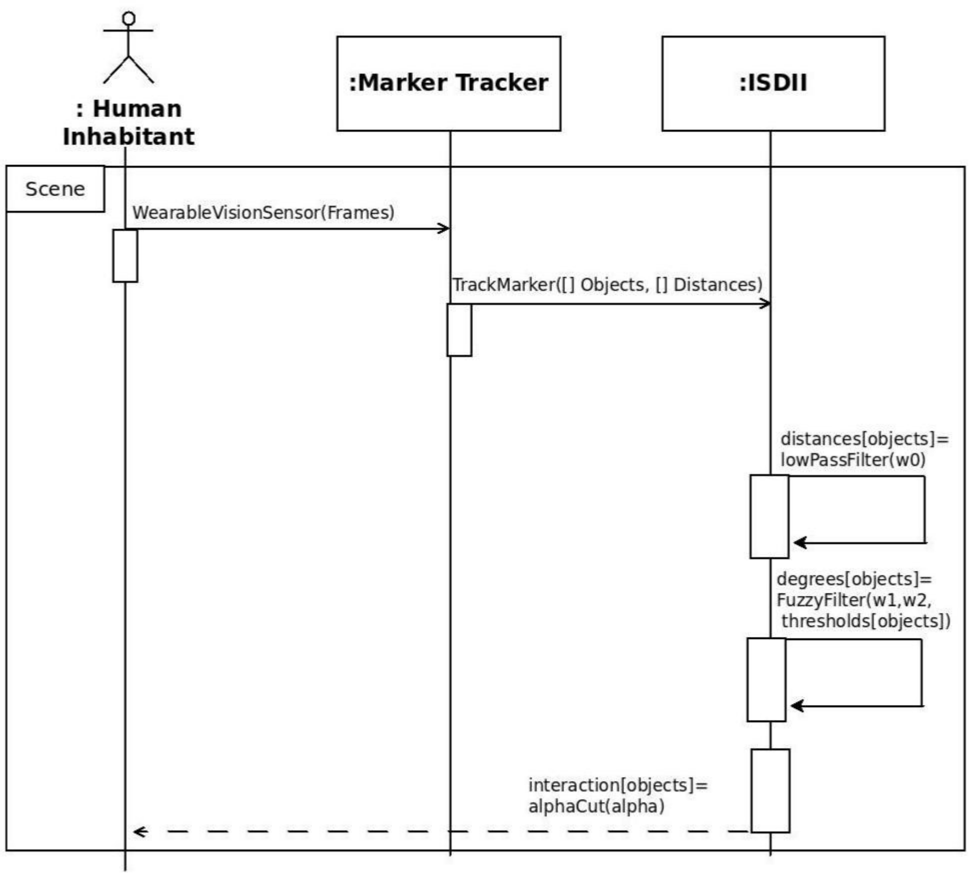


Figure 5.5: Sequence diagram of detecting object interaction in real-time scenes.

In summary, ISDII offers a solution of determining if an occupant is physically interacting with an object (*e.g.* opening a cupboard *vs.* if the marker was simply detected as the occupant navigated throughout the environment. ISDII achieves this through collecting the data stream of the occupant interacting with objects within the environment to allow a threshold distance to be determined for each object of interest. This threshold is the distance that the occupant is located in relation to the object of interest within the video stream. If an occupant is within this threshold it is determined they are interacting with the detected object.

This was achieved by collecting a number of recordings of an occupant interacting with objects of interest throughout the environment. This allowed a human expert to review the recordings and determine a threshold value based on the distance reported within the video stream. The next stage was to implement a low-pass filter to remove high-frequency noise, this allowed any unwanted artifacts (grainy/fuzzy areas) within the image to be removed which results in the an image that is clearer. A second filter was then developed to help reduce unwanted FPs within the video stream which resulted from general gaze activity within the environment. When an object is detected within the video stream the distance of the occupant from the marker is evaluated. An upper and lower threshold was set to determine at what point an object interaction has begun (lower threshold) and at which point an object interaction can be certain to have taken place (upper threshold). A sliding scale between the two thresholds then determines the confidence that an occupant-object interaction is being carried out, represented by a 0 for a lack of confidence (lower threshold) and 1 for total confidence that an interaction is being carried out (upper threshold).

5.2.2 Detection Algorithm

In this Section three scenarios are analysed of an occupant who wore Google glasses within a smart lab environment[3]. A series of markers were applied to objects within a smart lab and the researcher was instructed to enter the environment and proceed to complete pre-defined activities, while wearing a pair of Google Glasses. The three activities were: 1) making hot chocolate; 2) preparing a hot snack and; 3) washing dishes. A sequential breakdown of the objects interacted with during the completion of each activity is presented in Table 5.1.

To facilitate the experiments, a total of 18 markers (9 unique), were placed within the environment on the following objects: kitchen door, cupboard doors, a microwave, a refrigerator, a tap, and a chair. Multiple lighting conditions were

[3]https://drive.google.com/file/d/0B_rp8F6H7iwDNFVsUGpxQ1RqeDg/view?usp=sharing

Table 5.1: Breakdown of each activity and its corresponding object interactions.

1) Hot Chocolate	2) Hot Snack	3) Washing Dishes
Kitchen door	Kitchen door	Kitchen door
Cup cupboard	Fridge	Tap
Fridge	Plate cupboard	Cup cupboard
Microwave	Microwave	Cutlery cupboard
Tea/hot chocolate cupboard	Cutlery cupboard	Tea/hot chocolate cupboard
Cutlery cupboard	Microwave	Plate cupboard
Microwave	Chair	Kitchen door
Tea/hot chocolate cupboard	Kitchen door	N/A
Kitchen door	N/A	N/A

simulated via the use of window blinds and artificial lighting to provide a realistic context to the scenarios. Each scene was represented by the total number of frames, the duration of the scene and the percentage of frames during which an object was correctly identified (TP rate). An object was deemed correctly identified if the system reported the expected marker I.D. within the correct frame. *I.e.* if the fridge marker was expected the system was deemed to have correctly identified the object. As the fridge I.D. was reported that the marker was present, and the occupant within the distance thresholds as discussed the previous section. The percentage of correctly identified frames out of the total number of frames that an object was present was calculated to determine the detection ratio. The results are presented within Table 5.2 which displays the activity number, the total number of frames that comprised the video stream, the duration in seconds that a marker was within the camera FoV, and the total number of frames where a marker was present. It also presents the percentage of frames the respective algorithm successfully detected a marker out of the number of frames where a marker was present. The experiment was performed by a single participant who was a researcher and not representative of the target population.

Table 5.2: Breakdown of the duration and total number of video frames within each activity, along with the video duration and the number of frames the object was present. The detection ratio each algorithm achieved over the three activities is also presented.

Parameters				Detection Ratio	
Activity	Total Frames	Duration (s)	Object Frames	ArUco (%)	ORB (%)
1	2574	96	658	44.8	25.9
2	1567	52	624	44.8	22.7
3	1663	96	604	36.5	28.3

The order, duration, and interaction between the occupant and the objects varied across the three case scenarios. In addition, different lighting conditions were simulated during the scenarios to provide a realistic context to evaluate the algorithms. The marked objects in the smart lab were located in different positions in the room. Zenith lights provided varied lighting conditions when collecting the scenes[4]. The videos were recorded at 24fps and stored in MPEG-4 Part 14 (mp4) format conforming to Google Glass specifications.

There are a number of potential application areas this technology could have an impact on. Improving the accuracy of fiducial marker algorithms can be of valuable benefit to AR applications leading to more immersive experiences. There are also many applications within manufacturing as improvements in fiducial marker detection can aid in automatically detecting defects within the manufacturing process. This technology could also have an impact on assistive technologies with improved fiducial marker recognition leading to further development in assistive technologies, along with the wider healthcare industry such as medical image registration. The findings from the comparision of fiducial marker algorithms can have a number of potential contributions and as development of these algoritms continues to grow they will increase in value as a tool across a number of industries.

5.3 Results

As shown in Tables 5.3, 5.4, and 5.5, both algorithms provide improved performance in low blur and high brightness situations, with ArUco displaying a higher

[4] https://drive.google.com/file/d/0B_rp8F6H7iwDNFVsUGpxQ1RqeDg/view?usp=sharing

detection rate in general. The strength of ORB is its roubustness to various brightness conditions, as can be seen in Tables 5.3, 5.4 and 5.5 which shows that ORB has fewer instances were zero of the frames were detected compared to the ArUco algorithm. This is, in part, due to ORB's implementation of the Harris Corner Detection algorithm, which has been shown to have strong performance in low lighting conditions [74, 300]. An example of favourable and unfavourable conditions regarding movement and brightness are presented in Figure 5.6, brightness levels were categorised as follows, low – blinds closed and lights off, medium – blinds open and lights off, high – blinds open and lights on. In addition, the results from this evaluation provides the initial threshold distance references for ISDII to be adjusted by an expert.

Tables 5.3, 5.4 and 5.5 detail the objects sequentially interacted with during each scene, along with the average distance that each object was detected, the number of frames and duration of frames that the occupant-object interaction took place within. Tables 5.3 5.4 and 5.5 also specifies the lighting conditions during the interaction with each object, along with the calculated distance from the occupant's view point to the marker. Details of the simulated conditions are provided, specifying the amount of motion blur during the interaction and the level of ambient lighting. The detection ratio of ORB and ArUco algorithms are presented, displaying the proportion of frames where an object was detected within the duration window.

Both trackers provide their best performance in low motion and high brightness situations with ArUco being more accurate in general cases. The strength of ORB is viewed as the ability to accommodate low brightness conditions. The auto-focus of the wearable vision sensor proved to be critical for the marker tracker when the user or motion blur pixelates the frames disables the marker detection. This is a limitation of using Google Glass for these experiments, as a camera with a faster auto-focus may improve the results. An example of favourable and unfavourable situations of movement and brightness are shown in Figure 5.6.

Table 5.3: Results from activity one. Objects interacted with, average distance the interaction was determined to have taken place, the number of frames the interaction takes place within, the duration the interaction lasted, levels of brightness and motion blur. The percentage of frames were each algorithm correctly detected the object.

Objects	Output			Simulated Conditions		Detection Ratio	
Interaction Order	Avg. Distance (m)	# of Frames	Duration (s)	Brightness	Motion Blur	ArUco (%)	ORB (%)
Door	0.36	95	3.96	High	Normal	50.00	43.48
Cupboard A open	0.36	32	1.33	High	Low	78.79	50.00
Cupboard A closed	0.19	34	1.42	High	Low	61.29	66.67
Fridge open	0.29	47	1.96	High	High	56.25	21.28
Fridge closed	0.24	44	1.83	High	High	62.22	45.45
Microwave open	0.47	54	2.25	Low	High	3.64	0.00
Microwave closed	0.37	50	2.08	Low	High	17.65	6.00
Cupboard B open	0.22	38	1.58	Normal	Low	61.54	5.26
Cupboard B closed	0.30	49	2.04	Normal	Low	68.00	2.04
Cupboard C open	0	29	1.21	Low	High	0.00	31.02
Cupboard C closed	0	26	1.08	Low	High	0.00	11.54
Microwave open	0.44	42	1.75	Low	Normal	13.95	4.76
Microwave closed	0.37	24	1.00	Low	Normal	24.00	5.88
Cupboard D open	0.31	29	1.21	High	Low	80.00	10.34
Cupboard D closed	0.19	35	1.45	High	Low	69.44	2.86
Door open	0.20	125	5.21	Normal	High	44.44	23.33

Table 5.4: Results from activity two. Objects interacted with, average distance the interaction was determined to have taken place, the number of frames the interaction takes place within, the duration the interaction lasted, levels of brightness and motion blur. The percentage of frames were each algorithm correctly detected the object.

Objects	Output			Simulated Conditions		Detection Ratio	
Interaction Order	Avg. Distance (m)	# of Frames	Duration (s)	Brightness	Motion Blur	ArUco (%)	ORB (%)
Door	0.35	95	3.96	High	Normal	52.08	24.21
Tap on	0.32	101	4.28	Low	Low	39.22	3.79
Cupboard C open	0.21	41	1.71	High	Low	4.76	14.63
Cupboard C closed	0.00	24	1.00	High	Low	0.00	12.5
Cupboard A open	0.23	32	1.33	High	Low	78.79	81.08
Cupboard A closed	0.18	34	1.42	High	Low	80.00	87.50
Cupboard B open	0.32	54	2.25	Normal	Low	70.91	5.55
Cupboard B closed	0.20	35	1.46	Normal	Low	66.67	2.85
Cupboard D open	0.25	45	1.88	Normal	Low	43.48	48.88
Cupboard D closed	0.22	35	1.46	Normal	Low	58.33	60.00
Door open	0.32	111	4.79	Normal	High	25.42	7.82

Table 5.5: Results from activity three. Objects interacted with, average distance the interaction was determined to have taken place, the number of frames the interaction takes place within, the duration the interaction lasted, levels of brightness and motion blur. The percentage of frames were each algorithm correctly detected the object.

Objects	Output			Simulated Conditions		Detection Ratio	
Interaction Order	Avg. Distance (m)	# of Frames	Duration (s)	Brightness	Motion Blur	ArUco (%)	ORB (%)
Door open	0.28	58	2.42	High	Normal	72.80	22.41
Fridge open	0.30	48	2.00	High	High	79.59	45.83
Fridge closed	0.19	29	1.21	High	High	60.00	44.82
Cupboard D open	0.24	29	1.21	Normal	Low	13.33	87.50
Cupboard D closed	0.18	27	1.13	Normal	Low	60.71	88.88
Microwave open	0.25	35	1.46	Low	Normal	22.22	11.42
Microwave closed	0.23	50	2.08	Low	Normal	5.88	0.00
Cupboard C open	0.00	30	1.25	High	Low	0.00	46.15
Cupboard C closed	0.00	26	1.08	High	Low	0.00	50.00
Chair	0.35	159	6.63	Normal	Normal	40.00	13.20
Door open	0.25	111	4.63	Normal	high	22.32	21.52

A) B)

Figure 5.6: Frames from the wearable vision sensor showing first person view of interactions with objects. A) Low brightness and high motion blur situation. B) High brightness and low motion blur situation.

As discussed, an initial threshold value for objects was generated. These values were then adjusted by an expert to determine at what distance an occupant is determined to be interacting with an object. Table 5.6 details the average distance of detection as determined by ISDII as well as the final threshold distance after being modified by a human expert for each object.

The precision and recall have been evaluated from the ISDII output against the time window determined by an expert. An interaction had been determined when the interaction degree exceeds $\alpha - cut = 0.95$. The evaluation has included the full range of options for estimating the $\omega_0 \epsilon [0,1], \omega_1 \epsilon [0,5], \omega_2 \epsilon [0,5], \omega_1 < \omega_2$ with a step offset of 0.5. Table 5.7 presents the best precision results from the three scenes in function of $\omega_0, \omega_1, \omega_2$ and Table 5.8 displaying the best results for recall. The Fβ results are presented in Table 5.9.

The results are presented in Tables 5.7 and 5.8. While the precision results obtained by ISDII to determine actual interactions are promising, it relies on the accuracy of detections from the marker detection algorithm in order to return an improved recall. The lack of detections resulted in a low recall which cannot be improved through the filtering and estimation process presented. The Averaged Ratio Detection (ARD) from the detection algorithm in each scene must have matched the distance threshold value to be able to analyse the recall obtained by ISDII. This improved the ratio of marker detection due to the exponential smoothing filter. The averaged parameters have been set to allow a comparison of ISDII interaction estimations to expert-defined interaction estimations. The results in Figure 5.7, 5.8, and 5.9 presents the human expert defined degree of interaction along with an overlay of the ISDII defined interaction.

Finally, the averaged parameters were set to tune ISDII comparing the human-defined interactions with the estimation of ISDII in the three scenes.

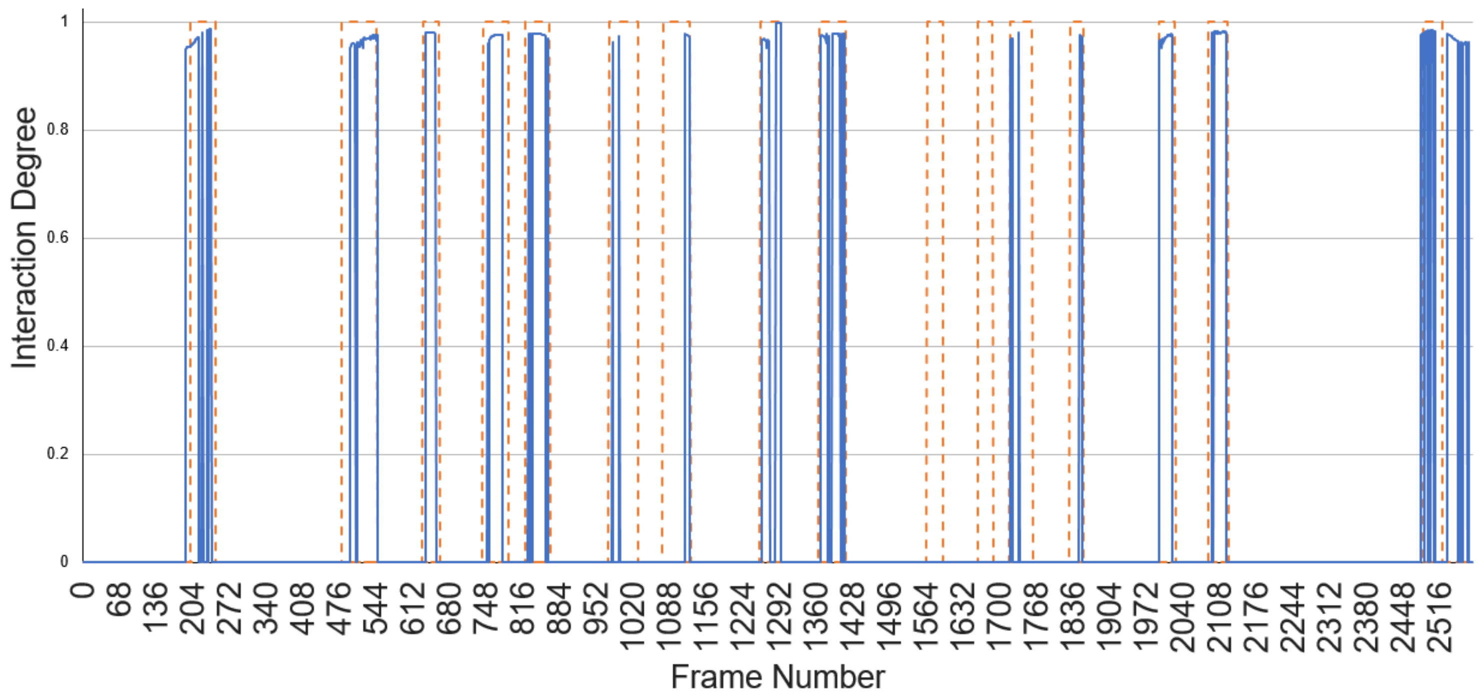


Figure 5.7: Comparison of ISDII *vs.* human-defined interactions for activity one. The orange dashed line representing the object interaction period as defined by a human expert. The blue solid line representing the object interaction period as estimated by ISDII.

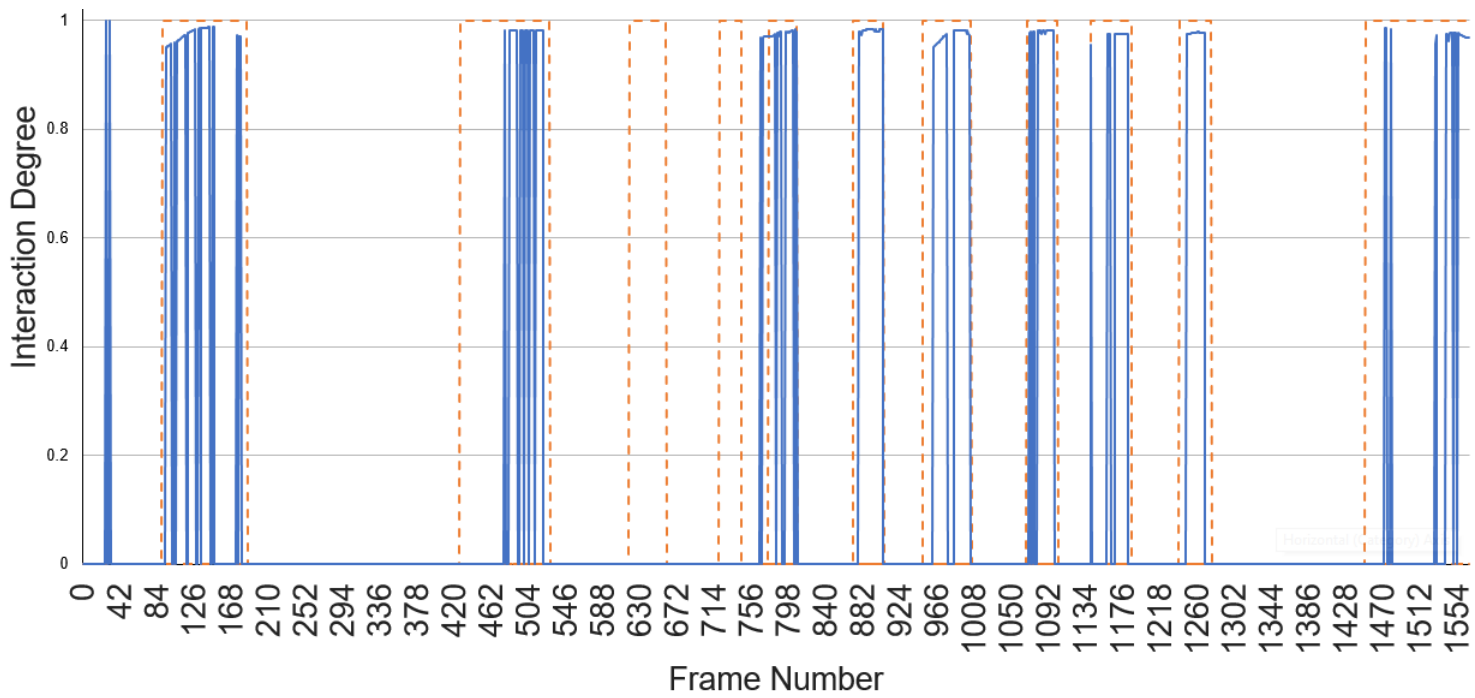


Figure 5.8: Comparison of ISDII *vs.* human-defined interactions for activity two. The orange dashed line representing the object interaction period as defined by a human expert. The blue solid line representing the object interaction period as estimated by ISDII.

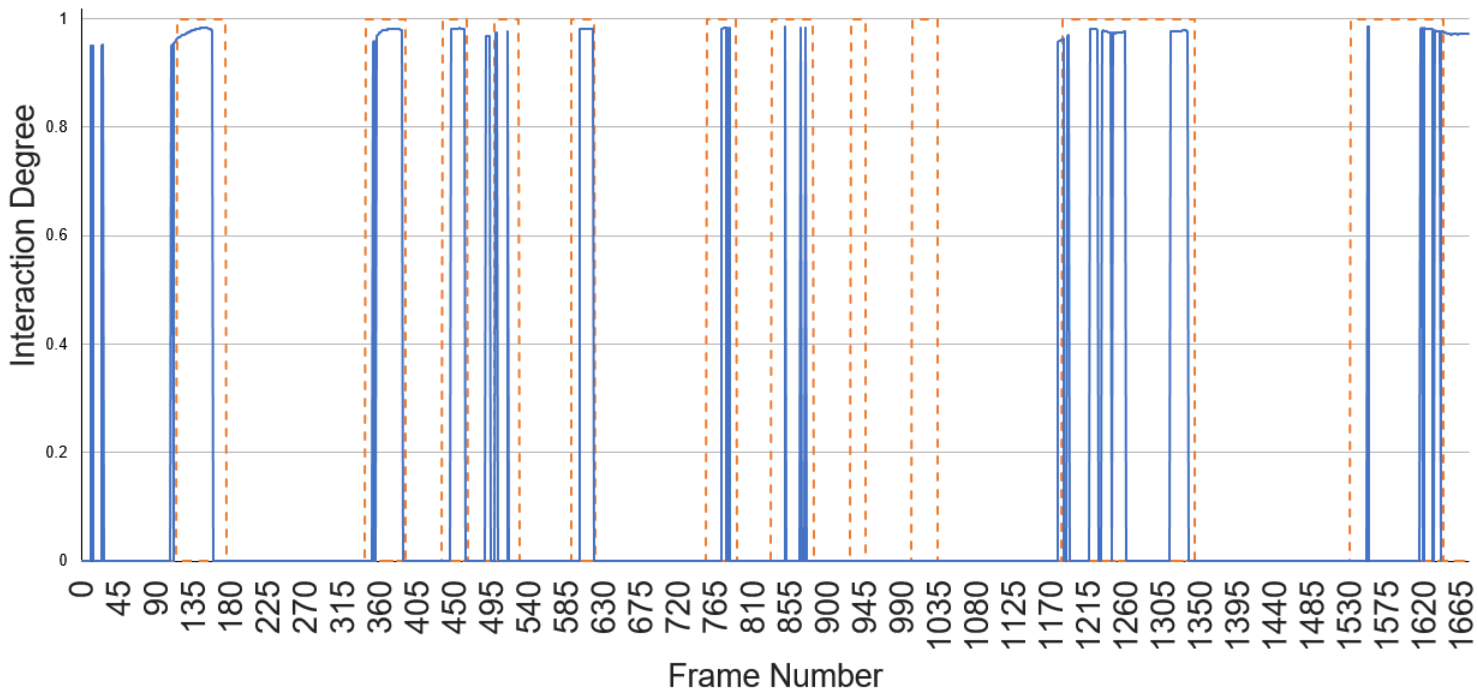


Figure 5.9: Comparison of ISDII *vs.* human-defined interactions for activity three. The orange dashed line representing the object interaction period as defined by a human expert. The blue solid line representing the object interaction period as estimated by ISDII.

Table 5.6: Threshold distances to objects.

Object	Average Distance (m)	Final Threshold Distance (m)
Chair	0.350	0.350
Cupboard A	0.240	0.235
Cupboard B	0.260	0.250
Cupboard C	0.240	0.250
Cupboard D	0.230	0.235
Door	0.296	0.300
Microwave	0.355	0.355
Fridge	0.255	0.255
Tap	0.320	0.320

Table 5.7: Best precision from scenes in function of $\omega_0; \omega_1; \omega_2$

Scene	Precision	$\omega_0; \omega_1; \omega_2$
1	1.00	[0.95;0.00;0.05]
2	0.98	[0.95;0.00;0.80]
3	1.00	[0.95;0.00;0.60]

Adjusting the threshold of object interaction offered improved performance when the detection algorithm provided an improved rate of detection, as the lack of detections shown in some scenes results in a loss of occupant-object interactions reported from ISDII. The final values of $\omega_0; \omega_1; \omega_2$ provided the best averaged parameters in all scenes, and resulted in a low computational overhead method of determining object interaction, as well as a method of isolating FP.

5.4 Discussion

The contributions offered by this chapter include the comparison of two popular off-the-shelf algorithms for feature detection in an AAL scenario. It also presents how lighting effects the performance of these two algorithms as well as that of motion blur, these are two very important factors when assessing the effectiveness

Table 5.8: Best recall from scenes in function of $\omega_0; \omega_1; \omega_2$

Scene	Recall	ARD	Value*/ARDl	$\omega_0; \omega_1; \omega_2$
1	0.45	0.43	1.05	[0.95;0.20;4.90]
2	0.45	0.47	0.95	[0.95;0.00;2.40]
3	0.37	0.34	1.09	[0.95;0.00;3.10]

Table 5.9: Best Fβ from scenes in function of $\omega_0; \omega_1; \omega_2$

Scene	Fβ	$\omega_0; \omega_1; \omega_2$
1	0.51	[0.95;0.20;2.20]
2	0.52	[0.95;0.00;2.45]
3	0.43	[0.95;0.00;1.65]
Average	0.49	[0.95;0.00;2.10]

of vision based aids and their feasibility in being applied to a real world situation.

ISDII is another contribution that this chapter has made. This itself has two contributions within. Namely, the implementation of a two stage filter which allows uncertainty in real time video based application to be reduced through exponential smoothing to reduce high frequency noise, and the second stage which involves the removal of isolated detections, such as those experienced through natural gaze activity. This used fuzzy logic to estimate the level of interaction the occupant is having with the object through distance estimation.

A final contribution from this chapter is the development of a system that does not require any user interaction in order to ensure that the best image angle is being captured. This challenge was previously identified and presented in Chapter Two as a limitation of existing systems. Occlusions that may be created through environmental objects, such as doors and large items of furniture, or occlusions generated by the occupant themselves, such as hands/head/torso occluding objects that they are interacting with [314]. This coupled with being a superior solution for object interaction due to the added advantages a head-mounted camera provides. Firstly, occlusions of the manipulated object tend to be lessened as the object being interacted with is usually the centre of attention for the occupant [314]. As the object is the centre of the occupant's attention the object is usually in the centre of the image and in focus, providing a high quality image for processing

[314]. Due to the high levels of noise that are typically present in egocentric videos many FP are unavoidable [315]. It can be difficult to identify the correct object as it is possible that multiple objects can be within the occupant's field of view. This is due to some areas of the environment being densely populated with relevant objects, such as the kitchen. Firstly, the ease with which it can be deployed within differing environments, the use of fiducial markers with an associated ID negates the need for specific training to each environment. This is due to the markers being associated with common static items that are commonly found within home environments, with the ID of the object being tied to the marker rather than any features of the object itself. Secondly, the use of a moving camera coupled with static objects reduces the issues traditionally seen with a static camera solution such as the limited field of view, which may require the installation of multiple cameras within an environment.

5.5 Conclusion

The results show that the ArUco algorithm is generally more accurate, with the ORB algorithm providing better performance in extreme light conditions. Based on the information from marker trackers, this chapter proposed an ISDII, which determines if the interaction is a TP by employing two filters: a low-pass filter and a fuzzy filter. A study was conducted to determine the performance of ISDII, showing an improved precision by reducing the number of detected FP. However, it is highly sensitive to FN from the detection algorithm which can result in a deteriorated recall result.

The proposed findings offers a non-intrusive method of detecting occupant object interaction and localisation. The use of a single head-worn camera provides a unique first person view of the environment and their activities, offering additional opportunities within the domain. The use of a first person camera also alleviates the need for the occupant to interact with the camera. As the camera is mounted within glasses, the field of view of the camera is more optimised for the direction that the object of interest for the occupant is positioned. This solution also minimises the cost in terms of hardware, implementation, and maintenance costs associated with alternative solutions, for example, dense sensor placement or static camera approaches. Given the target user group will be of an advanced age which typically have a lower level of technological ability this offers a key advantage. The user does not have to interact with the system or consider the position/placement of the wearable camera reducing user error and improving the quality of the data collected.

One limitation is that the system relies on a static marker within the environment, this is due to the system utilising the static markers to determine the location of the occupant. The scenario of a moving marker coupled with a moving camera would not allow the occupant's location to be determined within the environment. Additionally, due to the increased movement within the video stream due to the marker and camera moving independently the number of FP and FN would increase due to the additional movement blur within the stream. It should also be noted that a limitation is a lack of comparision against other algorithms designed specifically for fiducial marker detection other than ArUco. This can limit the generalisability of the findings and future work should include a further comparision against fiducial marker algorithms, such as Vuforia [316], AprilTags [317], and ARTag [318]. Additional limitations include the custom designed markers potentially not being as well optimised as the algorithmically design ArUco markers which could lead to the results favouring the ArUco algorithm. The placement of the markers within this study could introduce a bias due to multiple markers being applied to each object there is potential for markers to not be placed in the optimal location or they may suffer from issues such as occlusion or the camera viewing at a more extreme viewing angle. Lastly, the presented method is sensitive to the accuracy of the detection algorithm which can result in additional FNs which in turn will result in a lower recall value which can result in relevant information being missed. Chapter 6 investigates utilising Dempster-Shafer theory as a method of mitigating the loss of relevant information when determining which ADL an occupant is undertaking.

5.6 Associated Publications

Shewell, C, Medina-Quero, J, Espinilla, M, Nugent, C, Donnelly, M & Wang, HHY 2016, "Comparison of Fiducial Marker Detection and Object Interaction in Activities of Daily Living Utilising a Wearable Vision Sensor", International Journal of Communication Systems. https://doi.org/10.1002/dac.3223

Chapter 6: Activity Detection Incorporating Evidential Reasoning

6.1 Introduction

Activity recognition within Smart Environments is a key area of functionality as the ability of an occupant to carry out ADLs is a important metric to determine whether an occupant is able to continue living independently or if they need an increased level of support in order to remain within their own home. The accuracy of activity recognition within smart environments will always be subject to the reliability and validity of the sensors themselves. This is in part due to errors within the sensors which may report incorrect information or may miss sensor events completely thus leaving blanks within the data stream. While there has been some attempt among the research community to incorporate fuzzy logic within activity recognition such as Neural Networks [222], Dynamic Bayesian Networks [319], and Hidden Markov Models [279], Dempster-Shafer (DS) theory aims to handle the uncertainty introduced through the sensor errors in the smart environment. DS theory can provide improved results via increased reliability when compared to the previously discussed methods through its reasoning mechanism [320, 321]

The previous Chapters focused on the detection and filtering of fiducial markers in order to determine the occupant's location within an environment by detecting object interactions and the determination if these interactions were a TP through the use of the ISDII filter. Chapter 4 detailed a method of determining the occupant's location within an indoor environment through the use of a first person wearable camera and fiducial markers that were placed on key objects of interest within the environment. Chapter 5 built on this system in order to further filter out FPs that were detected either through mis-detecting a fiducial marker or through the detection of additional fiducial markers due to the occupant navigating throughout the environment or through general gaze activity. This was achieved through further filtering of the video stream, along with the creation of the ISDII system to detect the distance of the object from the occupant and whether that distance falls within pre-defined thresholds to determine if the occupant-object interaction is a TP. This Chapter discusses the implementation of a DS methodology which was previously developed by Hong *et al.* [12] to determine the probability of an activity having been carried out based on the detection of the occupant-object interactions as discussed in previous chapters. The hypothesis being considered is that does the application of DS theory improve the ability to recognise the user's activity within an environment.

6.2 Methodology

In order to apply a probability of the activity an initial belief value for each of the machine-vision detections needs to be established. From the results of the previous studies presented in Chapter 4 a belief value of 0.82 was determined for the machine-vision events (the detection algorithm successfully identified 143 instances out of a total of 175). This has been taken from the number of correctly identified objects within the total number expected within the video stream. The datasets that were presented in Chapter 3 have been used in order to demonstrate how DS theory can be implemented in order to recognise ADLs.

In order to determine the probability of an activity being carried out, a separate system was developed using Java which would accept the output from the machine-vision system as described in previous chapters. The machine-vision outputs the detected object as a String containing the object name and the time-stamp that the detection occurred. This is then accepted as input arguments within the DS component of the system. Figure 6.1 presents the flow of data through the system from the initial object detection, through the multiple stage filtering and the stages of assigning a probability belief to the activity. The seperation of the video streaming component from the DS component allows for a modular system. This allows the video to be streamed to multiple servers for processing should this be required, it can also allow a live video stream to be sent to family members should the need arise.

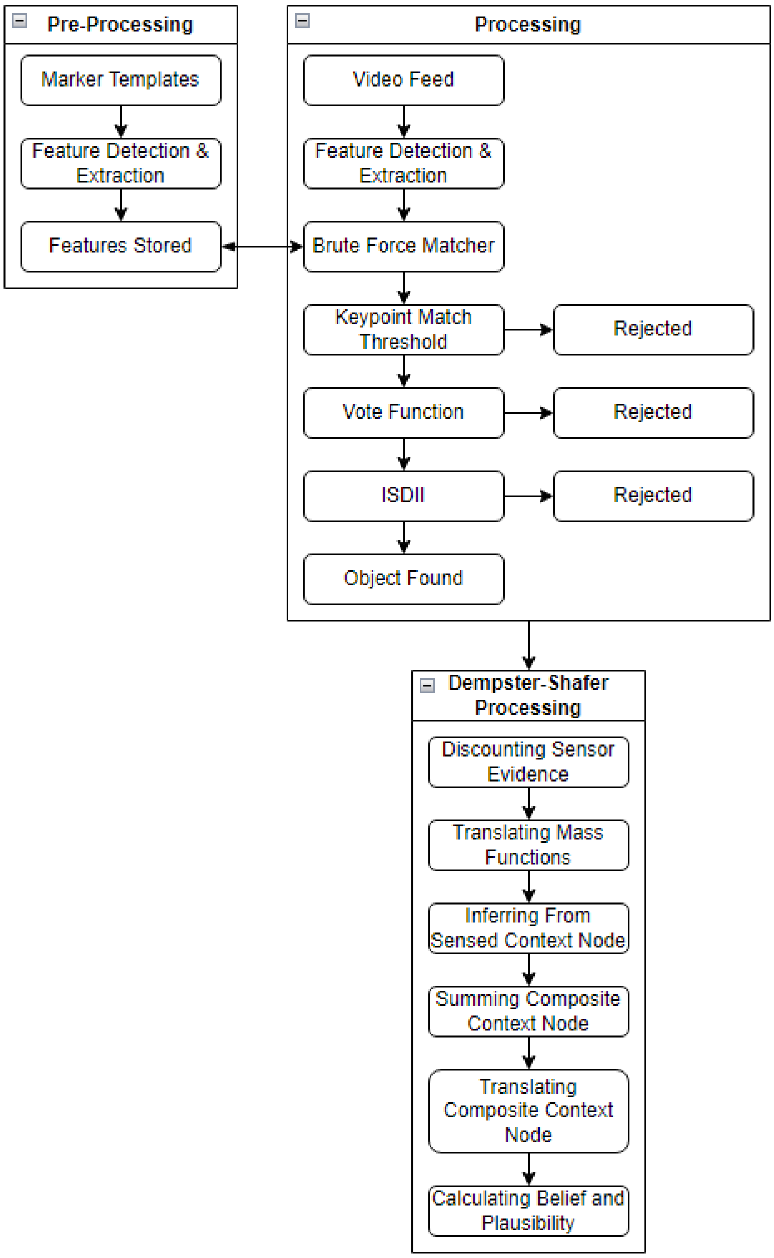


Figure 6.1: Flow of data through the system, from the object detection through to assigning probability to an activity belief.

6.2.1 Dempster-Shafer Theory

DS theory was originated by Dempster and was later formalised by Shafer [322]. It is a numerical uncertainty reasoning mechanism. Within their framework a problem is defined by a finite set of mutually exclusive hypotheses which form the frame of discernment (Θ).

A simple example may be considered is that of a door sensor. This sensor can be defined as either being opened, closed or static. Using this simple example then activated, static becomes a complete set of the possible door states; known as the frame of discernment for the door. It is then possible to numerically measure the belief on a single hypothesis, or a subset of hypotheses, by using a *mass function, m,* over the frame Θ, which satisfies the following conditions in equation 6.1 (where ϕ is the empty set) and 6.2 (where A is a subset of Θ), respectively.

$$m(\phi) = 0 \tag{6.1}$$

$$\sum_{A \subseteq \Theta} m(A) = 1 \tag{6.2}$$

The mass function is used to represent the distribution of a unit of belief over the frame, single elements, subsets or the whole set of the frame. When the door is closed or opened, a numerical representation can be applied by the mass function on the frame $\Theta = activated, static$ as $m(activated) = 1, m(static) = 0, m(activated, static) = 0$.
The occurrence of A is able to be inferred from the total mass of all the subsets of A, this is known as the *belief function* (Bel), as shown in equation 6.3

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{6.3}$$

$Bel(A)$ measures the degree of belief of information in support of A. Therefore, the likelihoods of hypotheses can be compared in order to determine the most likely hypothesis. DS Theory incorporates a range of probability values rather than a single probability, this is done in order to be able to represent uncertainty in the data. The Belief is the lower bounds of the probability with Plausibility being the upper bound, with Belief representing the degree to which the evidence supports A taking place and Plausibility representing the extent to which the evidence fails to refute that A that is taking place. Equation 6.4 presents how the Plausibility (Pls) is determined.

$$Pls(A) = \sum_{B \supseteq A} m(B) \tag{6.4}$$

Dempster's rule of combination provides a mechanism to compute the consensus of multiple independent sources of information which may be able to provide information about the same problem. N is the number of independent sources and m_i is the belief distribution given by the Nth source. Dempster's rule of combination then allows a new belief distribution which represents the consensus of N belief distributions as shown in equation 6.5.

$$m(C) = \frac{\sum\limits_{A \cap B = C} m_1(A)m_2(B)}{1 - \sum\limits_{A \cap B = \emptyset} m_1(A)m_2(B)} = \frac{\sum\limits_{A \cap B = C} m_1(A)m_2(B)}{1 - \sum\limits_{A \cap B \neq \emptyset} m_1(A)m_2(B)} \tag{6.5}$$

6.2.2 Case Study

In order to demonstrate how DS Theory is applied the following scenario will be stepped through. An occupant enters the kitchen via the kitchen door which is detected via an egocentric camera, the system then detects that the plate cupboard, fridge, and bread cupboard are interacted with. The chair is also interacted with, however, the fiducial marker is not successfully detected resulting in a missed sensor event. Events: kitchenDoor (TP), plateCupboard (TP), fridge (TP), chair (FN), and breadCupboard (TP). (Note: for readability the "cupboard" suffix will not be used within the worked example).

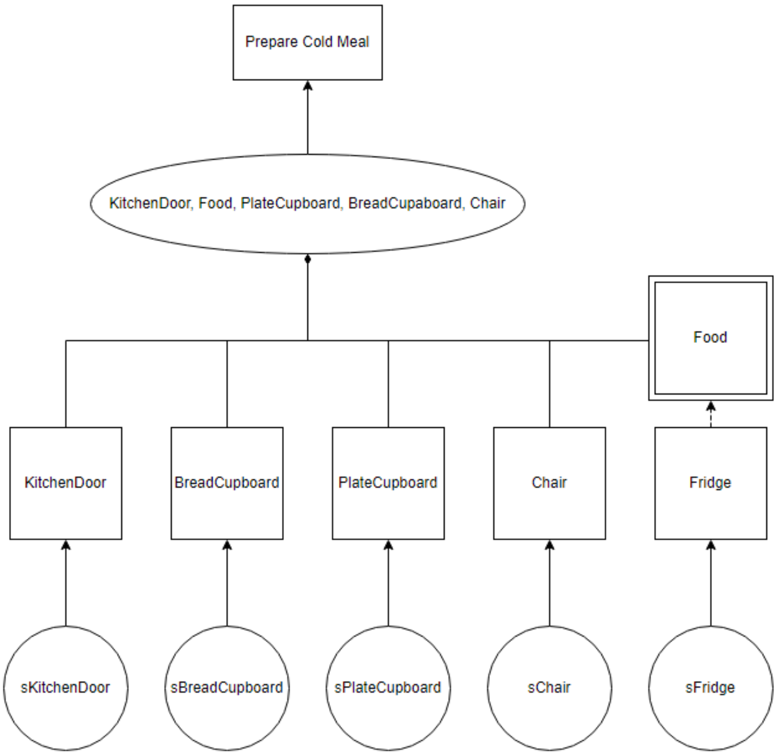


Figure 6.2: Example of multi-valued context for “Making a Cold Meal”..

Table 6.1: Summary of graphical notation used in Figure 6.2. [12]

Notation	Context
sensor	Sensor
Object	Object which is associated with a sensor.
Object	Object which has been derived from another object.
Composite Object	A composite object made up of multiple objects.
Activity	Activity

Step One: The discounted mass functions are calculated for each fiducial marker. Previous studies detailed in Chapters 4 shown an overall success rate of 82% for the vision based system, resulting in a discount rate of 18%. The discounted mass functions for each fiducial marker are presented below.

Table 6.2: Vision sensor belief values.

Object	Certainty	Uncertainty
Kitchen Door	0.82	0.18
Glass/Cup Cupboard	0.82	0.18
Microwave	0.82	0.18
Tea/Hot Chocolate Cupboard	0.82	0.18
Cutlery Cupboard	0.82	0.18
Fridge	0.82	0.18
Kettle	0.82	0.18
Bread Cupboard	0.82	0.18
Plate Cupboard	0.82	0.18
Tap	0.82	0.18
Living Room Door	0.82	0.18
Chair	0.82	0.18
Sofa	0.82	0.18
Telephone	0.82	0.18
TV	0.82	0.18

$$m^r_{skitchenDoor}(\{skitchenDoor\}) = 0.82,$$
$$m^r_{skitchenDoor}(\{skitchenDoor, \neg skitchenDoor\}) = 0.18;$$
$$m^r_{smicrowave}(\{\neg smicrowave\}) = 0.82,$$
$$m^r_{smicrowave}(\{smicrowave, \neg smicrowave\}) = 0.18;$$
$$m^r_{scutlery}(\{\neg scutlery\}) = 0.82,$$
$$m^r_{scutlery}(\{scutlery, \neg scutlery\}) = 0.18;$$
$$m^r_{skettle}(\{\neg skettle\}) = 0.82,$$
$$m^r_{skettle}(\{skettle, \neg skettle\}) = 0.18;$$
$$m^r_{stea/hot}(\{\neg stea/hot\}) = 0.82,$$
$$m^r_{stea/hot}(\{stea/hot, \neg stea/hot\}) = 0.18;$$
$$m^r_{sfridge}(\{sfridge\}) = 0.82,$$
$$m^r_{sfridge}(\{sfridge, \neg sfridge\}) = 0.18;$$
$$m^r_{stap}(\{\neg stap\}) = 0.82,$$
$$m^r_{stap}(\{stap, \neg stap\}) = 0.18;$$
$$m^r_{splates}(\{splates\}) = 0.82,$$
$$m^r_{splates}(\{splates, \neg splates\}) = 0.18;$$
$$m^r_{sbread}(\{sbread\}) = 0.82,$$
$$m^r_{sbread}(\{sbread, \neg sbread\}) = 0.18;$$
$$m^r_{slivingRoomDoor}(\{\neg slivingRoomDoor\}) = 0.82,$$
$$m^r_{slivingRoomDoor}(\{slivingRoomDoor, \neg slivingRoomDoor\}) = 0.18;$$
$$m^r_{schair}(\{\neg schair\}) = 0.82,$$
$$m^r_{schair}(\{schair, \neg schair\}) = 0.18;$$
$$m^r_{ssofa}(\{\neg ssofa\}) = 0.82,$$
$$m^r_{ssofa}(\{ssofa, \neg ssofa\}) = 0.18;$$
$$m^r_{stelephone}(\{\neg stelephone\}) = 0.82,$$
$$m^r_{stelephone}(\{stelephone, \neg stelephone\}) = 0.18;$$
$$m^r_{stelevision}(\{\neg stelevision\}) = 0.82,$$
$$m^r_{stelevision}(\{stelevision, \neg stelevision\}) = 0.18;$$
$$m^r_{sglass/cup}(\{\neg sglass/cup\}) = 0.82,$$
$$m^r_{sglass/cup}(\{sglass/cup, \neg sglass/cup\}) = 0.18;$$

Step Two: The mass functions are translated from the fiducial markers to the associated object. A detected fiducial marker indicates there has been an interaction with the associated object context. A fiducial marker along with the associated context maintains a compatible relationship that can be represented by multi-valued mapping as presented in Table 6.3. The mass functions calculated in Step One can then be translated to the associated object.

Table 6.3: Example of multi-valued mappings for "Making a Cold Meal".

Relationship	Multi-valued Mapping
sKitchenDoor ->kitchenDoor	{sKitchenDoor} ->{kitchenDoor}; {¬sKitchenDoor} ->{¬kitchenDoor}; {sKitchenDoor, ¬sKitchenDoor} ->{kitchenDoor, ¬kitchenDoor}.
kitchenDoor ->(kitchenDoor, bread, plates, chair, food)	{kitchenDoor} ->{(kitchenDoor, bread, plates, chair, food)}; {¬kitchenDoor} ->{¬(kitchenDoor, bread, plates, chair, food)}; {kitchenDoor, ¬kitchenDoor} ->{(kitchenDoor, bread, plates, chair, food), ¬(kitchenDoor, bread, plates, chair, food)}.
(kitchenDoor, bread, plates, chair, food) ->Prepare Cold Meal	{(kitchenDoor, bread, plates, chair, food)} ->{prepare cold meal}; {¬(kitchenDoor, bread, plates, chair, food)} ->{¬prepareColdMeal}; {(kitchenDoor, bread, plates, chair, food}, ¬(kitchenDoor, bread, plates, chair, food)} ->{parpareColdMeal ¬prepareColdMeal}.

$m_{kitchenDoor}(\{kitchenDoor\}) = m^r_{skitchenDoor}(\{skitchenDoor\}) = 0.82,$

$m_{kitchenDoor}(\{kitchenDoor, \neg kitchenDoor\}) =$

$m^r_{skitchenDoor}(\{skitchenDoor, \neg skitchenDoor\}) = 0.18;$

$m^r_{microwave}(\{\neg microwave\}) = m^r_{smicrowave}(\{\neg smicrowave\}) = 0.82,$

$m^r_{microwave}(\{microwave, \neg microwave\}) =$

$m^r_{smicrowave}(\{smicrowave, \neg smicrowave\}) = 0.18;$

$m^r_{cutlery}(\{\neg cutlery\}) = m^r_{scutlery}(\{\neg scutlery\}) = 0.82,$

$m^r_{cutlery}(\{cutlery, \neg cutlery\}) = m^r_{scutlery}(\{scutlery, \neg scutlery\}) = 0.18;$

$m^r_{kettle}(\{\neg kettle\}) = m^r_{skettle}(\{\neg skettle\}) = 0.82,$

$m^r_{kettle}(\{kettle, \neg kettle\}) = m^r_{skettle}(\{skettle, \neg skettle\}) = 0.18;$

$m^r_{tea/hot}(\{\neg tea/hot\}) = m^r_{stea/hot}(\{\neg stea/hot\}) = 0.82,$

$m^r_{tea/hot}(\{tea/hot, \neg tea/hot\}) = m^r_{stea/hot}(\{stea/hot, \neg stea/hot\}) = 0.18;$

$m_{fridge}(\{fridge\}) = m^r_{sfridge}(\{sfridge\}) = 0.82,$

$m_{fridge}(\{fridge, \neg fridge\}) = m^r_{sfridge}(\{sfridge, \neg sfridge\}) = 0.18;$

$m^r_{tap}(\{\neg tap\}) = m^r_{stap}(\{\neg stap\}) = 0.82,$

$m^r_{tap}(\{tap, \neg tap\}) = m^r_{stap}(\{stap, \neg stap\}) = 0.18;$

$m_{plates}(\{plates\}) = m^r_{splates}(\{splates\}) = 0.82,$

$m_{plates}(\{plates, \neg plates\}) = m^r_{splates}(\{splates, \neg splates\}) = 0.18;$

$m_{bread}(\{bread\}) = m^r_{sbread}(\{sbread\}) = 0.82,$

$m_{bread}(\{bread, \neg bread\}) =$

$m^r_{sbread}(\{sbread, \neg sbread\}) = 0.18;$

$m^r_{livingRoomDoor}(\{\neg livingRoomDoor\}) =$

$m^r_{slivingRoomDoor}(\{\neg slivingRoomDoor\}) = 0.82,$

$m^r_{livingRoomDoor}(\{livingRoomDoor, \neg livingRoomDoor\}) =$

$m^r_{slivingRoomDoor}(\{slivingRoomDoor, \neg slivingRoomDoor\}) = 0.18;$

$m^r_{chair}(\{\neg chair\}) = m^r_{schair}(\{\neg schair\}) = 0.82,$

$m^r_{chair}(\{chair, \neg chair\}) = m^r_{schair}(\{schair, \neg schair\}) = 0.18;$

$m^r_{sofa}(\{\neg sofa\}) = m^r_{ssofa}(\{\neg ssofa\}) = 0.82,$

$m^r_{sofa}(\{sofa, \neg sofa\}) = m^r_{ssofa}(\{ssofa, \neg ssofa\}) = 0.18;$

$m^r_{telephone}(\{\neg telephone\}) = m^r_{stelephone}(\{\neg stelephone\}) = 0.82,$

$m^r_{telephone}(\{telephone, \neg telephone\}) =$

$m^r_{stelephone}(\{stelephone, \neg stelephone\}) = 0.18;$

$m^r_{television}(\{\neg television\}) = m^r_{stelevision}(\{\neg stelevision\}) = 0.82,$

$m^r_{television}(\{television, \neg television\}) =$

$m^r_{stelevision}(\{stelevision, \neg stelevision\}) = 0.18;$

$m^r_{glass/cup}(\{\neg glass/cup\}) = m^r_{sglass/cup}(\{\neg sglass/cup\}) = 0.82,$

$m^r_{glass/cup}(\{glass/cup, \neg glass/cup\}) = m^r_{sglass/cup}(\{sglass/cup, \neg sglass/cup\}) = 0.18;$

Step Three: Inferring from a sensed context node to a deduced context node. In some cases contexts, such as “tea” and “hot chocolate”, are masked by their context, such as “tea/hot” in this case. The state of the context “tea/hot” can be detected by its associated fiducial marker. There is a heuristic relationship between between “tea/hot” and ‘tea” and/or “hot chocolate” and can be represented by an evidential mapping as presented in Table 6.4. The values presented in Table 6.4 were determined by examining the frequency with which each object was interacted with over the three routines. The mass functions of the deduced contexts can be calculated using the mass functions in Step Two and evidential mappings.

$m_{hotChocolate}(\{\neg hotChocolate\}) =$

$m_{tea/hot}(\{\neg tea/hot\}) * m(\{\neg tea/hot\} \rightarrow \{\neg hotChocolate\})$

$0.82 * 1 = 0.82,$

$m_{hotChocolate}(\{hotChocolate, \neg hotChocolate\}) =$

$m_{tea/hot}(\{tea/hot, \neg tea/hot\}) * m(\{tea/hot, \neg tea/hot\} \rightarrow$

$\{hotChocolate, \neg hotChocolate\})$

$0.18 * 1 = 0.18;$

$m_{tea}(\{\neg tea\}) =$

$m_{tea/hot}(\{\neg tea/hot\}) * m(\{\neg tea/hot\} \rightarrow \{\neg tea\})$

$0.82 * 1 = 0.82,$

$m_{tea}(\{tea, \neg tea\}) =$

$m_{tea/hot}(\{tea/hot, \neg tea/hot\}) * m(\{tea/hot, \neg tea/hot\} \rightarrow \{tea, \neg tea\})$

$0.18 * 1 = 0.18.$

$m_{cup}(\{\neg cup\}) =$

$m_{glass/cup}(\{\neg glass/cup\}) * m(\{\neg glass/cup\} \rightarrow \{\neg cup\})$

$0.82 * 1 = 0.82,$

$m_{cup}(\{cup, \neg cup\}) =$

$m_{glass/cup}(\{glass/cup, \neg glass/cup\}) * m(\{glass/cup, \neg glass/cup\} \rightarrow \{cup, \neg cup\})$

$0.18 * 1 = 0.18.$

$m_{glass}(\{\neg glass\}) =$

$m_{glass/cup}(\{\neg glass/cup\}) * m(\{\neg glass/cup\} \rightarrow \{\neg glass\})$

$0.82 * 1 = 0.82;$

$m_{glass}(\{glass, \neg glass\}) =$

$0.18 * 1 = 0.18.$

$m_{milk}(\{milk\}) =$

$m_{fridge}(\{fridge\}) * m(\{fridge\} \rightarrow \{milk\})$

$0.82 * 0.67 = 0.549,$

$m_{milk}(\{milk, \neg milk\}) =$

$m_{fridge}(\{fridge\}) * m(\{fridge\} \rightarrow \{milk, \neg milk\}) + m_{fridge}(\{fridge, \neg fridge\})$

$* m(\{fridge, \neg fridge\}) \rightarrow \{milk, \neg milk\})$

$0.82 * 0.33 + 0.18 * 1 = 0.451;$

$m_{food}(\{food\}) =$

$m_{fridge}(\{fridge\}) * m(\{fridge\} \rightarrow \{food\})$

$0.82 * 0.33 = 0.271,$

$m_{food}(\{food, \neg food\}) =$

$m_{fridge}(\{fridge\}) * m(\{fridge\} \rightarrow \{food, \neg food\} + m_{fridge}(\{fridge, \neg fridge\})$

$* m(\{fridge, \neg fridge\} \rightarrow \{food, \neg food\})$

$0.82 * 0.67 + 0.18 * 1 = 0.729.$

Table 6.4: Evidential Mappings based on the historical frequency of occupant-object interactions.

Object	Evidential Mapping
Tea/Hot ->Hot Chocolate	{tea/hot} ->{({hotChocolate}, 0.5), ({hotChocolate, ¬hotChocolate}, 0.5)}; {¬tea/hot} ->{({¬hotChocolate}, 1.0}; {tea/hot, ¬tea/hot} ->{({hotChocolate, ¬hotChocolate}, 1.0)}.
Tea/Hot ->Tea	{tea/hot} ->{({tea}, 0.5), ({tea, ¬tea}, 0.5)}; {¬tea/hot} ->{({¬tea}, 1.0}; {tea/hot, ¬tea/hot} ->{({tea, ¬tea}, 1.0)}.
Cup/Glass ->Cup	{cup/glass} ->{({cup}, 0.389), ({cup, ¬cup}, 0.611)}; {¬cup/glass} ->{({¬cup}, 1.0}; {cup/glass, ¬cup/glass} ->{({cup, ¬cup}, 1.0)}.
Cup/Glass ->Glass	{cup/glass} ->{({glass}, 0.611), ({glass, ¬glass}, 0.389)}; {¬cup/glass} ->{({¬glass}, 1.0}; {cup/glass, ¬cup/glass} ->{({glass, ¬glass}, 1.0)}.
Fridge ->Food	{food/milk} ->{({food}, 0.33), ({food, ¬food}, 0.67)}; {¬food/milk} ->{({¬food}, 1.0}; {food/milk, ¬food/milk} ->{({food, ¬food}, 1.0)}.
Fridge ->Milk	{food/milk} ->{({milk}, 0.67), ({milk, ¬milk}, 0.33)}; {¬food/milk} ->{({¬milk}, 1.0}; {food/milk, ¬food/milk} ->{({milk, ¬milk}, 1.0}.

Table 6.5: Belief values of deduced context nodes when sensed context node is triggered.

Object	Certainty	Uncertainty
Hot Chocolate	0.41	0.59
Tea	0.41	0.59
Cup	0.319	0.681
Milk	0.549	0.451
Glass	0.501	0.499
Food	0.271	0.729

Table 6.6: Belief values of deduced context nodes when sensed context node is not triggered.

Object	Certainty	Uncertainty
Hot Chocolate	0.82	0.18
Tea	0.82	0.18
Cup	0.82	0.18
Milk	0.82	0.18
Glass	0.82	0.18
Food	0.82	0.18

Step Four: Translating from the core context node to the composite context node. The individual contexts are then grouped into a multi-valued mapping as presented in Figure 6.2 (notation is detailed in Table 6.1). "kitchenDoor, food, plates, breadCupboard, chair" is the composite of "kitchenDoor", "food", "plates", "breadCupboard", and "chair". Table 6.3 presents the multi-valued mapping groups.

"Prepare Glass of Water" individual contexts: kitchen door, glass, and tap. Table 6.7 presents the multi-valued mapping for "Prepare Glass of Water".

Table 6.7: Multi-valued mappings for “Prepare glass of water”.

(kitchenDoor, glass, tap) ->Prepare Glass of Water	{(kitchenDoor, glass, tap)} ->{prepare glass of water}; {¬(kitchenDoor, glass, tap)} ->{¬prepareGlassOfWater}; {(kitchenDoor, glass, tap}, ¬(kitchenDoor, glass, tap)} ->{prepareGlassOfWater ¬prepareGlassOfWater}.

$$m0_{kgt}(\{kgt\}) = m_{kitchenDoor}(\{kitchenDoor\}) = 0.82,$$
$$m0_{kgt}(\{kgt, \neg kgt\}) = m_{kitchenDoor}(\{kitchenDoor, \neg kitchenDoor\}) = 0.18;$$
$$m1_{kgt}(\{\neg kgt\}) = m_{glass}(\{\neg glass\}) = 0.82,$$
$$m1_{kgt}(\{kgt, \neg kgt\}) = m_{glass}(\{glass, \neg glass\}) = 0.18;$$
$$m2_{kgt}(\{\neg kgt\}) = m_{tap}(\{\neg tap\}) = 0.82,$$
$$m2_{kgt}(\{kgt, \neg kgt\}) = m_{tap}(\{tap, \neg tap\}) = 0.18.$$

“Prepare Cup of Tea” individual contexts: kitchen door, kettle, tea, and cup. Table 6.8 presents the multi-valued mapping for “Prepare Cup of Tea”.

Table 6.8: Multi-valued mappings for “Prepare cup of tea”.

(kitchenDoor, kettle, tea, cup) ->Prepare Cup of Tea	{(kitchenDoor, kettle, tea, cup)} ->{prepare cup of tea}; {¬(kitchenDoor, kettle, tea, cup)} ->{¬prepareCupOfTea}; {(kitchenDoor, kettle, tea, cup}, ¬(kitchenDoor, kettle, tea, cup)} ->{prepareCupOfTea ¬prepareCupOfTea}.

$m0_{kctk}(\{kctk\}) = m_{kitchenDoor}(\{kitchenDoor\}) = 0.82,$

$m0_{kctk}(\{kctk, \neg kctk\}) = m_{kitchenDoor}(\{kitchenDoor, \neg kitchenDoor\}) =$ 0.18;

$m1_{kctk}(\{kctk\}) = m_{kettle}\{(\neg kettle\}) = 0.82,$

$m1_{kctk}(\{kctk\}) = m_{kettle}(\{kettle, \neg kettle\}) = 0.18.$

$m2kctck(\{\neg kctk\}) = m_{tea}(\{\neg tea\}) = 0.82,$

$m2_{kctk}(\{kctk, \neg kctk\}) = m_{tea}(\{tea, \neg tea\}) = 0.18;$

$m3_{kctk}(\{\neg kctk\}) = m_{cup}(\{\neg cup\}) = 0.82$

$m3_{kctk}(\{kctk, \neg kctk\}) = m_{cup}(\{cup, \neg cup\}) = 0.18;$

"Prepare Hot Chocolate" individual contexts: kitchen door, microwave, cutlery, hot chocolate, cup, and milk. Table 6.9 presents the multi-valued mapping for "Prepare Hot Chocolate".

Table 6.9: Multi-valued mappings for "Prepare hot chocolate".

(kitchenDoor, microwave, cutlery, hot chocolate, cup, milk) ->Prepare Hot Chocolate	{(kitchenDoor, microwave, cutlery, hot chocolate, cup, milk)} ->{prepare hot chocolate}; {¬(kitchenDoor, microwave, cutlery, hot chocolate, cup, milk)} ->{¬prepareHotChocolate}; {(kitchenDoor, microwave, cutlery, hot chocolate, cup, milk}, ¬(kitchenDoor, microwave, cutlery, hot chocolate, cup, milk)} ->{prepareHotChocolate ¬prepareHotChocolate}.

$m0_{kmchcm}(\{kmchcm\}) = m_{kitchenDoor}(\{kitchenDoor\}) = 0.82,$

$m0_{kmchcm}(\{kmchcm, \neg kmchcm\}) = m_{kitchenDoor}(\{kitchenDoor, \neg kitchenDoor\})$

$= 0.18;$

$m1_{kmchcm}(\{\neg kmchcm\}) = m_{microwave}(\{\neg microwave\}) = 0.82,$

$m1_{kmchcm}(\{kmchcm, \neg kmchcm\}) = m_{microwave}(\{microwave, \neg microwave\}) = 0.18;$

$m2_{kmchcm}(\{\neg kmchcm\}) = m_{cutlery}(\{\neg cutlery\}) = 0.82,$

$m2_{kmchcm}(\{kmchcm, \neg kmchcm\}) = m_{cutlery}(\{cutlery, \neg cutlery\}) = 0.18;$

$m3_{kmchcm}(\{\neg kmchcm\}) = m_{hotChocolate}(\{\neg hotChocolate\}) = 0.82,$

$m3_{kmchcm}(\{kmchcm, \neg kmchcm\}) = m_{hotChocolate}(\{hotChocolate, \neg hotChocoalte\})$

$= 0.18;$

$m4_{kmchcm}(\{\neg kmchcm\}) = m_{cup}(\{\neg cup\}) = 0.82,$

$m4_{kmchcm}(\{kmchcm, \neg kmchcm\}) = m_{cup}(\{cup, \neg cup\}) = 0.18;$

$m5_{kmchcm}(\{kmchcm\}) = m_{milk}(\{milk\}) = 0.549,$

$m5_{kmchcm}(\{kmchcm, \neg kmchcm\}) = m_{milk}(\{milk, \neg milk\}) = 0.451.$

"Prepare Glass of Milk" individual contexts: kitchen door, glass, milk. Table 6.10 presents the multi-valued mapping for "Prepare Glass of Milk".

Table 6.10: Multi-valued mappings for "Prepare glass of milk".

(kitchenDoor, glass, milk) ->Prepare Glass of Milk	{(kitchenDoor, glass, milk)} ->{prepare glass of milk}; {¬(kitchenDoor, glass, milk)} ->{¬prepareGlassOfMilk}; {(kitchenDoor, glass, milk}, ¬(kitchenDoor, glass, milk)} ->{prepareGlassOfMilk ¬prepareGlassOfMilk}.

$m0_{kgm}(\{kgm\}) = m_{kitchenDoor}(\{kitchenDoor\}) = 0.82,$

$m0_{kgm}(\{kgm, \neg kgm\}) = m_{kitchenDoor}(\{kitchenDoor, \neg kitchenDoor\}) = 0.18;$

$m1_{kgm}(\{\neg kgm\}) = m_{glass}(\{\neg glass\}) = 0.82,$

$m1_{kgm}(\{kgm, \neg kgm\}) = m_{glass}(\{glass, \neg glass\}) = 0.18;$

$m2_{kgm}(\{kgm\}) = m_{milk}(\{milk\}) = 0.549,$

$m2_{kgm}(\{kgm, \neg kgm\}) = m_{milk}(\{milk, \neg milk\}) = 0.451.$

"Make/Receive Phone Call" individual contexts: living room door, telephone. Table 6.11 presents the multi-valued mapping for "Make/Receive Phone Call".

Table 6.11: Multi-valued mappings for "Make/Receive Phone Call".

(livingRoomDoor, telephone) ->Make/Receive Phone Call	{(livingRoomDoor, telephone)} ->{make/recieve phone call}; {¬(livingRoomDoor, telephone)} ->{¬make/receive phone call}; {(livingRoomDoor, telephone}, ¬(livingRoomDoor, telephone)} ->{make/ReceivePhoneCall ¬make/ReceivePhoneCall}.

$m0_{lt}(\{\neg lt\}) = m_{livingRoomDoor}(\{\neg livingRoomDoor\}) = 0.82,$

$m0_{lt}(\{lt, \neg let\}) = m_{livingRoomDoor}(\{livingRoomDoor, \neg livingRoomDoor\}) = 0.18;$

$m1_{lt}(\{\neg lt\}) = m_{telephone}(\{\neg telephone\}) = 0.82,$

$m1_{lt}(\{lt, \neg lt\}) = m_{telephone}(\{telephone, \neg telephone\}) = 0.18.$

"Prepare Cold Meal" individual contexts: kitchen door, food, plates, bread cupboard, chair. Table 6.12 presents the multi-valued mapping for "Prepare Cold Meal".

Table 6.12: Multi-valued mappings for "Prepare Cold Meal".

(kitchenDoor, food, plates, bread, chair) ->Prepare Cold Meal	{(kitchenDoor, food, plates, bread, chair)} ->{prepare cold meal};
	{¬(kitchenDoor, food, plates, bread, chair)} ->{¬prepareColdMeal}; {(kitchenDoor, food, plates, bread, chair}, ¬(kitchenDoor, food, plates, bread, chair)} ->{prepareColdMeal ¬prepareColdMeal}.

$$m0_{kfpbc}(\{kfpbc\}) = m_{kitchenDoor}(\{kitchenDoor\}) = 0.82,$$
$$m0_{kfpbc}(\{kfpbc, \neg kfpbc\}) = m_{kitchenDoor}(\{kitchenDoor, \neg kitchenDoor\}) = 0.18;$$
$$m1_{kfpbc}(\{kfpbc\}) = m_{food}(\{food\}) = 0.271,$$
$$m1_{kfpbc}(\{kfpbc, \neg kfpbc\}) = m_{food}(\{food, \neg food\}) = 0.729;$$
$$m2_{kfpbc}(\{kfpbc\}) = m_{plates}(\{plates\}) = 0.82,$$
$$m2_{kfpbc}(\{kfpbc, \neg kfpbc\}) = m_{plates}(\{plates, \neg plates\}) = 0.18;$$
$$m3_{kfpbc}(\{kfpbc\}) = m_{bread}(\{bread\}) = 0.82,$$
$$m3_{kfpbc}(\{kfpbc, \neg kfpbc\}) = m_{bread}(\{bread, \neg bread\}) = 0.18;$$
$$m4_{kfpbc}(\{\neg kfpbc\}) = m_{chair}(\{\neg chair\}) = 0.82,$$
$$m4_{kfpbc}(\{kfpbc, \neg kfpbc\}) = m_{chair}(\{chair, \neg chair\}) = 0.18.$$

"Prepare Hot Meal" individual contexts: kitchen door, microwave, cutlery, food, plates, chair. Table 6.13 presents the multi-valued mapping for "Prepare Hot Meal".

Table 6.13: Multi-valued mappings for “Prepare Hot Meal”.

(kitchenDoor, microwave, cutlery, food, plates, chair) ->Prepare Hot Meal	{(kitchenDoor, microwave, cutlery, food, plates, chair)} ->{prepare hot meal};
	{¬(kitchenDoor, microwave, cutlery, food, plates, chair)} ->{¬prepareHotMeal}; {(kitchenDoor, microwave, cutlery, food, plates, chair}, ¬(kitchenDoor, microwave, cutlery, food, plates, chair)} ->{prepareHotMeal ¬prepareHotMeal}.

$m0_{kmcfpc}(\{kmcfpc\}) = m_{kitchenDoor}(\{kitchenDoor\}) = 0.82,$
$m0_{kmcfpc}(\{kmcfpc, \neg kmcfpc\}) = m_{kitchenDoor}(\{kitchenDoor, \neg kitchenDoor\})$
$= 0.18;$
$m1_{kmcfpc}(\{\neg kmcfpc\}) = m_{microwave}(\{\neg microwave\}) = 0.82,$
$m1_{kmcfpc}(\{kmcfpc, \neg kmcfpc\}) = m_{microwave}(\{microwave, \neg microwave\}) = 0.18;$
$m2_{kmcfpc}(\{\neg kmcfpc\}) = m_{cutlery}(\{\neg cutlery\}) = 0.82,$
$m2_{kmcfpc}(\{kmcfpc, \neg kmcfpc\}) = m_{cutlery}(\{cutlery, \neg cutlery\}) = 0.18;$
$m3_{kmcfpc}(\{kmcfpc\}) = m_{food}(\{food\}) = 0.271,$
$m3_{kmcfpc}(\{kmcfpc, \neg kmcfpc\}) = m_{food}(\{food, \neg food\}) = 0.729;$
$m4_{kmcfpc}(\{kmcfpc\}) = m_{plates}(\{plates\}) = 0.82,$
$m4_{kmcfpc}(\{kmcfpc, \neg kmcfpc\}) = m_{plates}(\{plates, \neg plates\}) = 0.18;$
$m5_{kmcfpc}(\{\neg kmcfpc\}) = m_{chair}(\{\neg chair\}) = 0.82,$
$m5_{kmcfpc}(\{kmcfpc, \neg kmcfpc\}) = m_{chair}(\{chair, \neg chair\}) - 0.18.$

“Watch TV” individual contexts: living room door, sofa, TV. Table 6.14 presents the multi-valued mapping for “Watch TV”.

Table 6.14: Multi-valued mappings for "Watch TV".

(livingRoomDoor, sofa, tv) ->Watch TV	{(livingRoomDoor, sofa, tv)} ->{watch tv};
	{¬(livingRoomDoor, sofa, tv)} ->{¬watchTV}; {(livingRoomDoor, sofa, tv}, ¬(livingRoomDoor, sofa, tv)} ->{watchTV ¬watchTV}.

$m0_{lst}(\{\neg lst\}) = m_{livingRoomDoor}(\{\neg livingRoomDoor\}) = 0.82$,

$m0_{lst}(\{lst, \neg lst\}) = m_{livingRoomDoor}(\{livingRoomDoor, \neg livingRoomDoor\}) = 0.18$;

$m1_{lst}(\{\neg lst\}) = m_{sofa}(\{\neg sofa\}) = 0.82$,

$m1_{lst}(\{lst, \neg lst\}) = m_{sofa}(\{sofa, \neg sofa\}) = 0.18$;

$m2_{lst}(\{\neg lst\}) = m_{tv}(\{\neg tv\}) = 0.82$,

$m2_{lst}(\{lst, \neg lst\}) = m_{tv}(\{tv, \neg tv\}) = 0.18$.

"Washing Dishes" individual contexts: kitchen door, cutlery, tap, plates, glass/cup cupboard. Table 6.15 presents the multi-valued mapping for "Washing Dishes".

Table 6.15: Multi-valued mappings for "Washing Dishes".

(kitchenDoor, cutlery, tap, plates, glass/cup) ->Washing Dishes	{(kitchenDoor, cutlery, tap, plates, glass/cup)} ->{washing dishes};
	{¬(kitchenDoor, cutlery, tap, plates, glass/cup)} ->{¬washingDishes}; {(kitchenDoor, cutlery, tap, plates, glass/cup}, ¬(kitchenDoor, cutlery, tap, plates, glass/cup)} ->{washingDishes ¬washingDishes}.

$m0_{kctpg/c}(\{kctpg/c\}) = m_{kitchenDoor}(\{kitchenDoor\}) = 0.82,$

$m0_{kctpg/c}(\{kctpg/c, \neg kctpg/c\}) = m_{kitchenDoor}(\{kitchenDoor, \neg kitchenDoor\})$

$= 0.18;$

$m1_{kctpg/c}(\{\neg kctpg/c\}) = m_{cutlery}(\{\neg cutlery\}) = 0.82,$

$m1_{kctpg/c}(\{kctpg/c, \neg kctpg/c\}) = m_{cutlery}(\{cutlery, \neg cutlery\}) = 0.18;$

$m2_{kctpg/c}(\{\neg kctpg/c\}) = m_{tap}(\{\neg tap\}) = 0.82,$

$m2_{kctpg/c}(\{kctpg/c, \neg kctpg/c\}) = m_{tap}(\{tap, \neg tap\}) = 0.18;$

$m3_{kctpg/c}(\{kctpg/c\}) = m_{plates}(\{plates\}) = 0.82,$

$m3_{kctpg/c}(\{kctpg/c, \neg kctpg/c\}) = m_{plates}(\{plates, \neg plates\}) = 0.18;$

$m4_{kctpg/c}(\{\neg kctpg/c\}) = m_{glass/cup}(\{\neg glass/cup\}) = 0.82,$

$m4_{kctpg/c}(\{kctpg/c, \neg kctpg/c\}) = m_{glass/cup}(\{glass/cup, \neg glass/cup\}) = 0.18.$

Step Five: Summing up a composite context node. For each multi-valued mapping in Step Four the mass functions are summed via an equally weighted sum operator so that the individual mass functions for "kitchenDoor", "food", "plates", "bread" and "chair" becomes "kitchenDoor, food, plates, bread, chair".

Composite node: kitchen door, glass, and tap.

$$\begin{aligned}
&m_{kgt}(\{kgt\})\\
&= 1/3(m0_{kgt} + m1_{kgt} + m2_{kgt})\\
&= 1/3(0.82 + 0 + 0) = 0.273,
\end{aligned}$$

$$\begin{aligned}
&m_{kgt}(\{\neg kgt\})\\
&= 1/3(m0_{kgt} + m1_{kgt} + m2_{kgt})\\
&= 1/3(0 + 0.82 + 0.82) = 0.547,
\end{aligned}$$

$$\begin{aligned}
&m_{kgt}(\{kgt, \neg kgt\})\\
&= 1/3(m0_{kgt} + m1_{kgt} + m2_{kgt})\\
&= 1/3(0.18 + 0.18 + 0.18) = 0.18.
\end{aligned}$$

Composite node: kitchen door, kettle, tea, and cup.

$$m_{kctk}(\{kctk\})$$
$$= 1/4(m0_{kctk} + m1_{kctk} + m2_{kctk} + m3_{kctk})$$
$$= 1/4(0.82 + 0 + 0 + 0) = 0.205,$$

$$m_{kctk}(\{\neg kctk\})$$
$$1/4(m0_{kctk} + m1_{kctk} + m2kctk + m3)kctk$$
$$1/4(0 + 0.82 + 0.82 + 0.82) = 0.615,$$

$$m_{kctk}(\{kctk, \neg kctk\})$$
$$1/4(m0_{kctk} + m1_{kctk} + m2_{kctk} + m3_{kctk})$$
$$1/4(0.18 + 0.18 + 0.18 + 0.18) = 0.18.$$

Composite node: kitchen door, microwave, cutlery, hot chocolate, cup, and milk.

$$m_{kmchcm}(\{kmchcm\})$$
$$= 1/6(m0_{kmchcm} + m1_{kmchcm} + m2_{kmchcm} + m3_{kmchcm} + m4_{kmchcm} + m5_{kmchcm})$$
$$= 1/6(0.82 + 0 + 00 + 0 + 0.549) = 0.228,$$

$$m_{kmchcm}(\{\neg kmchcm\})$$
$$= 1/6(m0_{kmchcm} + m1_{kmchcm} + m2_{kmchcm} + m3_{kmchcm} + m4_{kmchcm} + m5_{kmchcm})$$
$$= 1/6(0 + 0.82 + 0.82 + 0.82 + 0.82 + 0) = 0.547,$$

$$m_{kmchcm}(\{kmchcm, \neg kmchcm\})$$
$$1/6(m0_{kmchcm} + m1_{kmchcm} + m2_{kmchcm} + m3_{kmchcm} + m4_{kmchcm} + m5_{kmchcm})$$
$$1/6(0.18 + 0.18 + 0.18 + 0.18 + 0.18 + 0.451) = 0.225.$$

Composite node: kitchen door, glass, and milk.

$$m_{kgm}(\{lgm\})$$
$$= 1/3(m0_{kgm} + m1_{kgm} + m2_{kgm})$$
$$= 1/3(0.82 + 0 + 0.549) = 0.456,$$

$$m_{kgm}(\{\neg kgm\})$$
$$= 1/3(m0_{kgm} + m1_{kgm} + m2_{kgm})$$
$$= 1/3(0 + 0.82 + 0) = 0.273,$$

$$m_{kgm}(\{kgm, \neg kgm\})$$
$$1/3(m0_{kgm} + m1_{kgm} + m2_{kgm})$$
$$1/3(0.18 + 0.18 + 0.451) = 0.270.$$

Composite node: living room door and telephone.

$$m_{lt}(\{lt\})$$
$$1/2(m0_{lt} + m1_{lt})$$
$$1/2(0 + 0) = 0,$$

$$m_{it}(\{it\})$$
$$= 1/2(m0_{lt} + m1_{lt})$$
$$= 1/2(0.82 + 0.82 = 0.82,$$

$$m_{lt}(\{lt, \neg lt\})$$
$$= 1/2(m0_{lt} + m1_{lt})$$
$$1/2(0.18 + 0.18) = 0.18.$$

Composite node: kitchen door, food, plates, bread, and chair.

$$
\begin{aligned}
&m_{kfpbc}(\{kfpbc\})\\
&= 1/5(m0_{kfpbc} + m1_{kfpbc} + m2_{kfpbc} + m3_{kfpbc} + m4_{kfpbc})\\
&= 1/5(0.82 + 0.271 + 0.82 + 0.82 + 0) = 0.546,
\end{aligned}
$$

$$
\begin{aligned}
&m_{kfpbc}(\{\neg kfpbc\})\\
&= 1/5(m0_{kfpbc} + m1_{kfpbc} + m2_{kfpbc} + m3_{kfpbc} + m4_{kfpbc})\\
&= 1/5(0 + 0 + 0 + 0 + 0.82) = 0.164,
\end{aligned}
$$

$$
\begin{aligned}
&m_{kfpbc}(\{kfpbc, \neg kfpbc\})\\
&= 1/5(m0_{kfpbc} + m1_{kfpbc} + m2_{kfpbc} + m3_{kfpbc} + m4_{kfpbc}\\
&= 1/5(0.18 + 0.729 + 0.18 + 0.18 + 0.18) = 0.29.
\end{aligned}
$$

Composite node: kitchen door, microwave, cutlery, food, plates, and chair.

$$
\begin{aligned}
&m_{kmcfpc}(\{kmcfpc\})\\
&= 1/6(m0_{kmcfpc} + m1_{kmcfpc} + m2_{kmcfpc} + m3_{kmcfpc} + m4_{kmcfpc} + m5_{kmcfpc})\\
&= 1/6(0.82 + 0 + 0 + 0.271 + 0.82 + 0) = 0.319,
\end{aligned}
$$

$$
\begin{aligned}
&m_{kmcfpc}(\{\neg kmcfpc\})\\
&= 1/6(m0_{kmcfpc} + m1_{kmcfpc} + m2_{kmcfpc} + m2_{kmcfpc} + m3_{kmcfpc} + m4_{kmcfpc}\\
&+ m5_{kmcfpc})\\
&= 1/6(0 + 0.82 + 0.82 + 0 + 0 + 0.82) = 0.41,
\end{aligned}
$$

$$
\begin{aligned}
&m_{kmcfpc}(\{kmcfpc, \neg kmcfpc\})\\
&= 1/6(m0_{kmcfpc} + m1_{kmcfpc} + m2_{kmcfpc} + m3_{kmcfpc} + m4_{kmcfpc} + m5_{kmcfpc})\\
&= 1/6(0.18 + 0.18 + 0.18 + 0.729 + 0.18 + 0.18) = 0.271.
\end{aligned}
$$

Composite node: living room door, sofa, and TV.

$$
\begin{aligned}
& m_{lst}(\{lst\}) \\
& = 1/3(m0_{lst} + m1_{lst} + m2_{lst}) \\
& 1/3(0 + 0 + 0) = 0,
\end{aligned}
$$

$$
\begin{aligned}
& m_{lst}(\{\neg lst\}) \\
& = 1/3(m0_{lst} + m1_{lst} + m2_{lst}) \\
& = 1/3(0.82 + 0.82 + 0.82) = 0.82,
\end{aligned}
$$

$$
\begin{aligned}
& m_{lst}(\{lst, \neg lst\}) \\
& = 1/3(0.18 + 0.18 + 0.18) = 0.18.
\end{aligned}
$$

Composite node: kitchen door, cutlery, tap, plates, and glass/cup.

$$
\begin{aligned}
& m_{kctpg/c}(\{kctpg/c\}) \\
& 1/5(m0_{kctpg/c} + m1_{kctpg/c} + m2_{kctpg/c} + m3_{kctpg/c} + m4_{kctpg/c}) \\
& = 1/5(0.82 + 0 + 0 + 0.82 + 0) = 0.328,
\end{aligned}
$$

$$
\begin{aligned}
& m_{kctpg/c}(\{\neg kctpg/c\}) \\
& = 1/5(m0_{kctpg/c} + m1_{kctpg/c} + m2_{kctpg/c} + m3_{kctpg/c} + m4_{kctpg/c}) \\
& = 1/5(0 + 0.82 + 0.82 + 0 + 0.82) = 0.492,
\end{aligned}
$$

$$
\begin{aligned}
& m_{kctpg/c}(\{kctpg/c, \neg kctpg/c\}) \\
& = 1/5(m0_{kctpg/c} + m1_{kctpg/c} + m2_{kctpg/c} + m3_{kctpg/c} + m4_{kctpg/c} \\
& = 1/5(0.18 + 0.18 + 0.18 + 0.18 + 0.18) = 0.18.
\end{aligned}
$$

Step Six: Translating from a composite context node to an activity. Each mass function from the multi-valued mappings, as shown in Table 6.7 – 6.15 can be translated to an activity, *e.g.* “kitchenDoor, food, plates, bread, chair” can therefore be mapped to “Prepare Cold Meal”.

Prepare Water: kitchen door, glass, and tap.

$$m_{prepareWater}(\{prepareWater\}) = m_{kgt}(\{kgt\}) = 0.273,$$
$$m_{prepareWater}(\{\neg prepareWater\}) = m_{kgt})\{\neg kgt\}) = 0.547,$$
$$m_{prepareWater}(\{prepareWater, \neg prepareWater\}) = m_{kgt}(\{kgt, \neg kgt\}) = 0.18.$$

Prepare Tea: kitchen door, kettle, tea, and cup.

$$m_{prepareTea}(\{prepareTea\}) = m_{kctk}(\{kctk\}) = 0.205,$$
$$m_{prepareTea}(\{\neg prepareTea\}) = m_{kctk}(\{\neg kctk\}) = 0.615,$$
$$m_{prepareTea}(\{prepareTea, \neg prepareTea\}) = m_{kctk}(\{kctk, \neg kctk\}) = 0.18.$$

Prepare Hot Chocolate: kitchen door, microwave, cutlery, hot chocolate, cup, and milk.

$$m_{prepareHotChocolate}(\{prepareHotChocolate\}) = m_{kmchcm}(\{kmchcm\}) = 0.228,$$
$$m_{prepareHotChocolate}(\{\neg prepareHotChocolate\}) = m_{kmchcm}(\{\neg kmchcm\}) = 0.547,$$
$$m_{prepareHotChocolate}(\{prepareHotChocolate, \neg prepareHotChocolate\})$$
$$= m_{kmchcm}(\{kmchcm, \neg kmchcm\}) = 0.225.$$

Drink Milk: kithcen door, glass, and milk.

$$m_{drinkMilk}(\{drinkMilk\}) = m_{kgm}(\{kgm\}) = 0.456,$$
$$m_{drinkMilk}(\{\neg drinkMilk\}) = m_{kgm}(\{\neg kgm\}) = 0.273,$$
$$m_{drinkMilk}(\{drinkMilk, \neg drinkMilk\}) = m_{kgm}(\{kgm, \neg kgm\}) = 0.270.$$

Phone call: living room door and telephone.

$$
\begin{aligned}
&m_{phoneCall}(\{phoneCall\}) = m_{lt}(\{lt\}) = 0,\\
&m_{phoneCall}(\{\neg phoneCall\}) = m_{lt}(\{\neg lt\}) = 0.82,\\
&m_{phoneCall}(\{phoneCall, \neg phoneCall\}) = m_{lt}(\{lt, \neg lt\}) = 0.18.
\end{aligned}
$$

Prepare cold meal: kitchen door, food, plates, bread, and chair.

$$
\begin{aligned}
&m_{prepareColdMeal}(\{prepareColdMeal\}) = m_{kfpbc}(\{kfpbc\}) = 0.546,\\
&m_{prepareColdMeal}(\{\neg prepareColdMeal\}) = m_{kfpbc}(\{\neg kfpbc\}) = 0.164,\\
&m_{prepareColdMeal}(\{prepareColdMeal, \neg prepareColdMeal\})\\
&= m_{kfpbc}(\{kfpbc, \neg kfpbc\}) = 0.29.
\end{aligned}
$$

Prepare hot meal: kitchen door, microwave, cutlery, food, plates, and chair.

$$
\begin{aligned}
&m_{prepareHotMeal}(\{prepareHotMeal\}) = m_{kmcfpc}(\{kmcfpc\}) = 0.319,\\
&m_{prepareHotMeal}(\{\neg prepareHotMeal\}) = m_{kmcfpc}(\{\neg kmcfpc\}) = 0.41,\\
&m_{prepareHotMeal}(\{prepareHotMeal, \neg prepareHotMeal\})\\
&= m_{kmcfpc}(\{kmcfpc, \neg kmcfpc\}) = 0.271.
\end{aligned}
$$

Watch TV: living room door, sofa, and TV.

$$
\begin{aligned}
&m_{watchTV}(\{watchTV\}) = m_{lst}(\{lst\}) = 0,\\
&m_{watchTV}(\{\neg watchTV\}) = m_{lst}(\{\neg lst\}) = 0.82,\\
&m_{watchTV}(\{watchTV, \neg watchTV\}) = m_{lst}(\{lst, \neg lst\}) = 0.18.
\end{aligned}
$$

Wash dishes: kitchen door, cutlery, tap, plates, and glass/cup.

$$m_{washDishes}(\{washDishes\}) = m_{kctpg/c}(\{kctpg/c\}) = 0.328,$$
$$m_{washDishes}(\{\neg washDishes\}) = m_{kctpg/c}(\{\neg kctpg/c\}) = 0.492,$$
$$m_{washDishes}(\{washDishes, \neg washDishes\})$$
$$= m_{kctpg/c}(\{kctpg/c, \neg kctpg/c\}) = 0.18.$$

Step Seven: Calculating belief and plausibility. The belief and plausibility can be calculated from the mass functions on each activity *e.g.* *"Prepare cold meal"*.

Prepare Water.

$$Bel(\{prepareWater\}) = m(\{prepareWater\}) = 0.273,$$

$$Pls(\{prepareWater\})$$
$$= m(\{prepareWater\}) + m(\{prepareWater, \neg prepareWater\})$$
$$= 0.273 + 0.18$$
$$= 0.453.$$

Prepare Tea.

$$Bel(\{prepareTea\}) = m(\{prepareTea\}) = 0.205,$$

$$Pls(\{prepareTea\})$$
$$= m(\{prepareTea\}) + m(\{prepareTea, \neg prepareTea\})$$
$$= 0.205 + 0.18$$
$$= 0.385.$$

Prepare Hot Chocolate.

$$Bel(\{prepareHotChocolate\}) = m(\{prepareHotChocolate\}) = 0.228,$$

$$\begin{aligned}
&Pls(\{prepareHotChocolate\})\\
&= m(\{prepareHotChocolate\})+\\
&m(\{prepareHotChocolate, \neg prepareHotChocolate\})\\
&= 0.228 + 0.225\\
&= 0.453.
\end{aligned}$$

Drink Milk.

$$Bel(\{drinkMilk\}) = m(\{drinkMilk\}) = 0.456,$$

$$\begin{aligned}
&Pls(\{drinkMilk\})\\
&= m(\{drinkMilk\}) + m(\{drinkMilk, \neg drinkMilk\})\\
&= 0.456 + 0.270\\
&= 0.726.
\end{aligned}$$

Phone Call.

$$Bel(\{phoneCall\}) = m(\{phoneCall\}) = 0,$$

$$\begin{aligned}
&Pls(\{phoneCall\})\\
&= m(\{phoneCall\}) + m(\{phoneCall, \neg phoneCall;\})\\
&= 0 + 0.18\\
&= 0.18.
\end{aligned}$$

Prepare Cold Meal.

$$Bel(\{prepareColdMeal\}) = m(\{prepareColdMeal\}) = 0.546,$$

$$\begin{aligned}
&Pls(\{prepareColdMeal\})\\
&= m(\{prepareColdMeal\}) + m)\{prepareColdMeal, \neg prepareColdMeal\})\\
&= 0.546 + 0.29\\
&= 0.836.
\end{aligned}$$

Prepare Hot Meal.

$$Bel(\{prepareHotMeal\}) = m(\{prepareHotMeal\}) = 0.319,$$

$$\begin{aligned}
&Pls(\{prepareHotMeal\})\\
&= m(\{prepareHotMeal\}) + m(\{prepareHotMeal, \neg prepareHotMeal\})\\
&= 0.319 + 0.271\\
&= 0.59.
\end{aligned}$$

Watch TV.

$$Bel(\{watchTV\}) = m(\{watchTV\}) = 0,$$

$$\begin{aligned}
&Pls(\{watchTV\})\\
&= m(\{watchTV\}) + m(\{watchTV, \neg watchTV\})\\
&= 0 + 0.18\\
&= 0.18.
\end{aligned}$$

Wash Dishes.

$$Bel(\{washDishes\}) = m(\{washDishes\}) = 0.328,$$

$$\begin{aligned}
&Pls(\{washDishes\})\\
&= m(\{washDishes\}) + m(\{washDishes, \neg washDishes\})\\
&= 0.328 + 0.18\\
&= 0.508.
\end{aligned}$$

The belief and plausibility for each activity were then compared using the maximisation operator.

$Bel(activity)$

$= max(Bel(\{prepareWater\}), Bel(\{prepareTea\}), Bel(\{prepareHotChocolate\}),$

$Bel(\{drinkMilk\}), Bel(\{phoneCall\}), Bel(\{prepareColdMeal\}),$

$Bel(\{prepareHotMeal\}), Bel(\{watchTV\}), Bel(\{washDishes\}))$

$= max(0.273, 0.205, 0.228, 0.456, 0, 0.546, 0.319, 0, 0.328)$

$= 0.546$

$Pls(activity)$

$= max(Pls(\{prepareWater\}), Pls(\{prepareTea\}), Pls(\{prepareHotChocolate\}),$

$Pls(\{drinkMilk\}), Pls(\{phoneCall\}), Pls(\{prepareColdMeal\}),$

$Pls(\{prepareHotMeal\}), Pls(\{watchTV\}), Pls(\{washDishes\}))$

$= max(0.453, 0.385, 0.453, 0.726, 0.18, 0.836, 0.59, 0.18, 0.508)$

$= 0.836.$

We can then therefore be confident that "Prepare Cold Meal" has been carried out due to the resulting belief and plausibility values being higher than that of alternative activities. With "Prepare Cold Meal" having a Belief of 0.546 and a Plausibility of 0.836 *vs.* the next most likely activity of "Drink Milk" with a Belief of 0.456 and a Plausibility of 0.726.

6.2.3 Experimental Routine

The experiment routine used in this Chapter is based on the Ulster and Jaèn datasets as discussed previously in Chapter Three. Each dataset consists of nine ADLs chosen to represent a wide range of activities that an occupant would carry out during their day to day routine. These are presented in Table 6.16. The Ulster dataset consists of 32 activities, as shown in Table 6.17, which were recordings of a live video stream to demonstrate how the system will perform in a real world scenario in which artifacts may be present in the video stream along with other potential issues, such as missing or corrupt video frames. The Jaèn dataset is made up of three routines consisting of ten ADLs for a total of 30 activities spread over three routines. These can be seen in Table 6.18. The system was then tested to establish if it could correctly estimated the activity that was being carried out through the use of DS Theory.

Table 6.16: The list of available activities and their corresponding activity number.

Activity Number	Activity
1	Prepare glass of water (PW)
2	Prepare cup of tea (PT)
3	Prepare hot chocolate (PHC)
4	Prepare glass of milk (PM)
5	Make/receive phone call (PC)
6	Prepare cold meal (PCM)
7	Prepare hot meal (PHM)
8	Watch TV (WTV)
9	Washing dishes (WD)

Table 6.17: The list of activities (represented by their associated activity number) that make up each routine from the Ulster dataset.

Routine One	Routine Two	Routine Three
3	4	3
1	6	1
7	1	5
9	5	7
8	1	1
1	2	8
8	8	2
6	7	8
9	9	6
1	8	9
N/A	1	4

Table 6.18: The list of activities (represented by their associated activity number) that make up each routine from the Jaèn dataset.

Routine One	Routine Two	Routine Three
3	4	3
1	6	1
7	1	5
9	5	7
8	1	1
1	2	8
8	8	2
6	7	8
9	9	6
1	8	9

To provide a benchmark, a traditional machine learning approach was also implemented on the dataset with each sensor acting as a feature within the model. Three well known supervised machine learning classification algorithms were implemented. In supervised classification, methods a set of training data is used to train the model with a separate dataset used to test the model accuracy. The dataset consisted of a total of 64 events consisting of 20 features. Due to the small nature of the dataset the dataset was split with 80% being supplied as the training set, and the remaining 20% being supplied as the unseen testing set. Cross validation was not used due to conerns of overfitting as K-fold cross validation has been shown to be strongly biased when applied to small datasets [323]. Additionally, due to the limited data, cross fold validation can potentially increase variance due to the similarity between the training and testing data reducing it's ability to generalise to further datasets. Cross fold validation can also result in minority classes being used frequently for testing which can lead to skewed results in the models performance. The three machine leaning algorithms used in this Chapter were:

Naive Bayes [324]: Naive Bayes is a probabilistic model which is based on the Bayes theorem [325] and is well known for multi-class prediction. Naive Bayes is based on probability models that have strong independence assumptions built in, *i.e.* the classifier assumes that each input variable is independent.
Random Forest [326]: Random Forest classifiers combine the output of multiple decision trees, utilising both bagging and feature randomness to create an uncorrelated forest of decision trees with the goal of reaching a final, single result.

Multilayer Perceptron (MLP) [327]: An MLP is a class of artificial neural network and an integral part of deep learning. MLP's are made up of three layers – an input layer which is responsible for receiving input from the dataset, one or more hidden layer(s) which is responsible for applying weights to the inputs and directing them though an activation function, and an output layer which is responsible for outputting a value or vector of values.

6.3 Results

This Section describes the results from both the machine learning benchmark results and the results from the DS component of the system. Table 6.20 displays the overall results from the three machine learning algorithms when applied to the collected dataset from both Ulster and Jaén labs in Chapter 3 as well as the overall results from the DS theory application. The confusion matrices are presented in Figure 6.3. Table 6.21 presents a further breakdown of the results displaying the recall, precision, and F-measure for the classifiers and DS theory. The recall score is a measure of how many instances that the system correctly predicted. Precision is a measure of how many of the predicted instances are correctly predicted. F-measure is the harmonic mean of recall and precision. The results from the detection algorithm from Chapter 4 and 5 were represented by binary data with a marker detection (TP and FP) represented by a one, non-detection represented as a zero, and missing/corrupt data represented as unknown "?". The data was then ran through three ML classifiers to determine if they could correctly identify the activity that was being undertaken by the occupant. It should be noted that as the Ulster dataset was streamed live from the Google Glass device it has a much higher rate of missed sensor events, this was mainly due to the Glass device reducing its processing speed to help with cooling, as detailed in Chapter Three – Section 3.3.2, which results in a higher rate of missed frames and corruption in the video stream.

Tables 6.22 – 6.27 presents the breakdown of the belief and plausibility values from DS theory along with the identified activity from each routine for the datasets from both Ulster and Jaén. Each table presents the activity that the researcher was carrying out as the "Expected Activity" with the determined activity from the DS implemention presented as the "Identified Activity". The belief and plausibility in the identified activity is also presented.

Table 6.20: Percentage of correctly and incorrectly classified results from the machine learning models and DS theory.

Method	Correctly Classified	Incorrectly Classified
Naive Bayes	66.7%	33.3%
RandomForest	75.0%	25.0%
Multilayer Perceptron	66.7%	33.3%
DS Theory	84.0%	16.0%

```
a b c d e f g h i   <-- classified as
4 0 0 0 0 0 0 0 0 | a = DrinkWater
0 0 0 0 0 0 0 0 0 | b = PrepareTea
0 0 0 0 0 0 0 0 0 | c = PrepareHotChocolate
0 0 0 0 0 0 0 0 1 | d = PrepareMilk
0 0 0 0 1 0 0 0 0 | e = PhoneCall
0 0 0 0 0 0 0 0 0 | f = PrepareColdMeal
0 0 0 3 0 0 1 0 0 | g = PrepareHotMeal
0 0 0 0 0 0 0 1 0 | h = WatchTV
0 0 0 0 0 0 0 0 1 | i = WashDishes
```

(a) Naive Bayes

```
a b c d e f g h i   <-- classified as
4 0 0 0 0 0 0 0 0 | a = DrinkWater
0 0 0 0 0 0 0 0 0 | b = PrepareTea
0 0 0 0 0 0 0 0 0 | c = PrepareHotChocolate
0 0 0 0 0 0 0 0 1 | d = PrepareMilk
0 0 0 0 1 0 0 0 0 | e = PhoneCall
0 0 0 0 0 0 0 0 0 | f = PrepareColdMeal
0 0 1 1 0 0 2 0 0 | g = PrepareHotMeal
0 0 0 0 0 0 0 1 0 | h = WatchTV
0 0 0 0 0 0 0 0 1 | i = WashDishes
```

(b) Random Forest

```
a b c d e f g h i   <-- classified as
4 0 0 0 0 0 0 0 0 | a = DrinkWater
0 0 0 0 0 0 0 0 0 | b = PrepareTea
0 0 0 0 0 0 0 0 0 | c = PrepareHotChocolate
1 0 0 0 0 0 0 0 0 | d = PrepareMilk
0 0 0 0 1 0 0 0 0 | e = PhoneCall
0 0 0 0 0 0 0 0 0 | f = PrepareColdMeal
0 0 0 2 1 0 1 0 0 | g = PrepareHotMeal
0 0 0 0 0 0 0 1 0 | h = WatchTV
0 0 0 0 0 0 0 0 1 | i = WashDishes
```

(c) Multiplayer Perceptron

Figure 6.3: Confusion matrices for the three ML classifiers.

Table 6.21: Recall, Precision, and F-measure scores for the three classifiers and DS theory.

Method	Recall	Precision	F-Measure
Naive Bayes	0.667	0.875	0.757
RandomForest	0.750	0.875	0.808
Multilayer Perceptron	0.667	0.808	0.731
DS Theory	1.000	0.840	0.920

Table 6.22: DS results from the Routine One from the Ulster dataset.

Expected Activity	Identified Activity	Belief	Plausibility
Drink Water	Drink Water	0.714	1.000
Drink Water	Drink Water	0.714	1.000
Drink Water	Drink Water	0.714	1.000
Prepare Hot Chocolate	Drink Milk	0.440	0.727
Cold Meal	Cold Meal	0.546	0.836
Hot Meal	Hot Meal	0.273	0.453
Watch TV	Watch TV	0.820	1.000
Watch TV	Watch TV	0.820	1.000
Wash Dishes	Wash Dishes	0.328	0.508
Wash Dishes	Drink Milk	0.440	0.727

Table 6.23: DS results from the Routine Two from the Ulster dataset.

Expected Activity	Identified Activity	Belief	Plausibility
Drink Water	Drink Water	0.440	0.727
Drink Water	Drink Water	0.714	1.000
Drink Water	Drink Water	0.714	1.000
Prepare Tea	Prepare Tea	0.490	0.795
Drink Milk	Drink Milk	0.623	1.000
Phone Call	Phone Call	0.820	1.000
Cold Meal	Cold Meal	0.328	0.508
Hot Meal	Hot Meal	0.410	0.590
Watch TV	Watch TV	0.820	1.000
Watch TV	Watch TV	0.820	1.000
Wash Dishes	Wash Dishes	0.820	1.000

Table 6.24: DS results from the Routine Three from the Ulster dataset.

Expected Activity	Identified Activity	Belief	Plausibility
Drink Water	Drink Water	0.714	1.000
Drink Water	Drink Water	0.714	1.000
Prepare Tea	Prepare Tea	0.490	0.795
Prepare Hot Chocolate	Prepare Hot Chocolate	0.478	0.727
Drink Milk	Drink Milk	0.623	1.000
Phone Call	Phone Call	0.820	1.000
Cold Meal	Cold Meal	0.546	0.836
Hot Meal	Hot Meal	0.547	0.727
Watch TV	Watch TV	0.820	0.727
Watch TV	Watch TV	0.820	1.000
Wash Dishes	Drink Water	0.714	1.000

Table 6.25: DS results from the Routine One from the Jaén dataset.

Expected Activity	Identified Activity	Belief	Plausibility
Prepare Hot Chocolate	Drink Milk	0.623	1.000
Drink Water	Drink Water	0.714	1.000
Hot Meal	Hot Meal	0.546	0.836
Wash Dishes	Wash Dishes	0.714	1.000
Drink Water	Drink Water	0.714	1.000
Watch TV	Watch TV	0.820	1.000
Cold Meal	Cold Meal	0.546	0.836
Wash Dishes	Drink Water	0.714	1.000
Drink Water	Drink Water	0.714	1.000
Watch TV	Watch TV	0.820	1.000

Table 6.26: DS results from the Routine Two from the Jaén dataset.

Expected Activity	Identified Activity	Belief	Plausibility
Drink Milk	Drink Milk	0.623	1.000
Cold Meal	Cold Meal	0.546	0.836
Drink Water	Drink Water	0.714	1.000
Phone Call	Phone Call	0.820	1.000
Drink Water	Drink Water	0.714	1.000
Prepare Tea	Prepare Tea	0.592	1.000
Watch TV	Watch TV	0.820	1.000
Hot Meal	Drink Milk	0.456	0.727
Wash Dishes	Drink Water	0.714	1.000
Watch TV	Watch TV	0.820	1.000
Drink Water	Drink Water	0.714	1.000

Table 6.27: DS results from the Routine Three from the Jaén dataset.

Expected Activity	Identified Activity	Belief	Plausibility
Prepare Hot Chocolate	Drink Milk	0.623	1.000
Drink Water	Drink Water	0.714	1.000
Phone Call	Phone Call	0.820	1.000
Hot Meal	Cold Meal	0.546	0.836
Drink Water	Drink Water	0.714	1.000
Watch TV	Watch TV	0.820	1.000
Prepare Tea	Prepare Tea	0.592	1.000
Watch TV	Watch TV	0.820	1.000
Cold Meal	Drink Milk	0.456	0.727
Drink Water	Drink Water	0.714	1.000
Drink Milk	Drink Milk	0.623	1.000

6.4 Discussion

As presented in Table 6.21, the DS implementation demonstrates an improved recall and F-Measure score over the traditional ML methods. An improved recall score demonstrates that the system misclassifies fewer activites than the ML approaches. This is of importance within the domain of AAL as the misclassification of an activity as an activity being misclassified as an alternative activity could cause confusion to the occupant along with a lack of timely and relevant support. The difference in performance is due to there being missing/corrupt data within the dataset. Due to DS theory being able to deal with uncertainty has resulted in the detection of activities that were misclassified by the ML algorithms. This is particularly evident with more complex tasks due to the higher likelihood of missing or corrupt sensor data. An example of this would be the activity "prepareHotMeal" which was misclassified by all three ML algorithms in the majority of cases.

However, the ML methods show an improved precision over the DS system. However, it should be noted that it is not possible to state if this is significantly significant or not. To determine statistical significance a statistical hypothesis test would need to be ran. Due to the comparison being comprised by multiple groups (DS theory and multiple ML algorithms) an Analysis of Variance (ANOVA) test would be suitable. In practical terms, the lower precision of the DS system can be beneficial when compared to the higher precision from the ML methods. This is due to a high precision potentially causing false alarms. If a system is overly precise it may trigger support or interventions for activities which may be classed as normal but that was not accurately identified by the system. This improved precision was due to an increased number of activities being recognised incorrectly as FPs due to the DS system being sensitive to certain activities having a low number of differentiating sensor profiles. An example of this would be the activities "Phone Call" and "Watch TV", with "Phone Call" consisting of the living room door and telephone sensors with "Watch TV" relying on the living room door and TV sensors. As a result if key sensor events are not detected they can have a large effect on the resulting belief and plausibility values. This can result in a reduced performance if an occupant's daily routine is made up with a number of activities with a low number of object interactions, due to not having enough evidence to accuratly differentiate between activities. A potential solution to mitigiate this challenge could be to introduce an additional sensor modality for low interaction activities to aid in differentiating the activities. For example, a contact sensor could be added to the phone and the TV remote to offer additional evidence of the activity being carried out.

One distinct advantage the DS system offers over an ML approach is that of a lack of training required when new activities are added or when the system is applied to a new environment. Should a new object/senor be added to the system it would only require the name of the object and the belief/disbelief values for the associated sensor. This offers a powerful advantage when coupled with the machine-vision aspect of the system which also does not require training, only requiring a template of the new fiducial marker along with an associated label. An additional advantage is the ability to apply the DS system to multiple environments without requiring retraining to the environment as demonstrated by implementing the system to both the Ulster and Jaén lab environments with no additional modification to the system required.

The contribution offered by this Chapter is an implementation of DS theory to that of an egocentric camera in order to correctly identify activities of daily living within a real world smart environment. This aids in alleviating the problem of unreliable sensor evidence [328, 329], particularly within a machine-vision system where a high number of variables, such as light, viewing angle, *etc.*, can effect the accuracy of an object recognition system [254]. This is corroborated by the results shown within this Chapter which shows a high level of accuracy maintained when performing activity recognition even when a number of sensor events are missed, though it should be noted that some missing values will have a greater effect on the results than others. This is due to some sensor events being key in accurately identifying the activity, such as Drink Water and Drink Milk which share sensor events with only one sensor event being unique in each case. Drink Water relies on the kitchen door, glass/cup cupboard, and tap sensor and Drink Milk relies on the kitchen door, glass/cup cupboard, and fridge. This small differentiation can result in the activity either not being detected or miss-classified as another similar activity.

There are a number of practical applications that this research can be applied to in the real world through offering a more robust method for object recognition. Thus offering a solution to the previously identified challenges of unreliable sensor evidence [328, 329]. Particularly in the real world where missed sensor events can be caused by issues such as faulty sensors which can effect the accuracy and reliability of activity recognition. Additionally, this study has highlighted the importance of certain sensor events being crucial for identifiying particular activities which can aid in informing the design and implementation of future AAL systems.

6.5 Conclusion

This Chapter presented a method of applying DS theory to a machine-vision based system in order to calculate a probabilistic belief of an ADL being carried out. A worked example has also been presented to demonstrate the concept and the system also showed how unreliable sensor evidence can be overcome to still provide an accurate estimation of the activity being carried out. The system was tested on a real world dataset which consisted of 64 activities over six routines from two separate smart-lab environments. The presented method displayed the ability to reliably detect the correct activity in a majority of cases with an overall percentage of correctly identified activities of approximately 84%. The proposed approach offers the advantage of detecting ADLs even with missing sensor values and offers increased reliability and safety with the domain of ambient assisted living and further moving towards a vision of 'aging in place'.

Furthermore, new activities can be added to the system without the need for existing data on the activities. This is due to the activities being made up of a composite of unique objects, this allowing a new activity to be added by including the composite objects in the activity template.

Chapter 7: Conclusion

7.1 Introduction

This Chapter will provide a reflection on the work presented within this thesis. This includes a discussion on the contributions to knowledge. The objectives detailed in Chapter 1 will be revisited to assess if the research carried out within this thesis has met these objectives. This will be presented along with discussion on the areas of future investigation as well as concluding remarks. This research aimed to investigate the use of machine-vision based approaches to support those at home who traditionally may require assistance to carry out their activities of daily living through the use of improved location accuracy and activity recognition via evidential reasoning. In order to achieve this aim a technical solution was developed leveraging an egocentric camera to detect fiducial markers that have been placed on key objects throughout the environment. To aid in improving the accuracy of marker detection a distance estimation tool was researched to estimate if an object interaction was genuine and was not caused through navigation of the environment. DS theory was then implemented to further increase the accuracy of detecting the occupant's activity when taking into account uncertainty within the data.

The research objectives which were identified in Chapter 1 are presented below, along with a discussion on how these objectives were met throughout this thesis.

7.2 Discussion of Objectives

As discussed in Chapter One, there has been a remarkable increase in life expectancy throughout the world [13]. This increase in the older section of society has resulted in an increase on the demand placed on the healthcare due to the large percentage of adults who require long-term support for independent living. In order to aid in alleviating this burden being placed on healthcare, researchers have investigated the use of smart home and wearable technology to develop approaches that will allow individuals to live within their own home for longer.

The aim of this thesis was to investigate the use of machine-vision based approaches to support those at home who may traditionally require assistance to carry out their ADLs through the use of improved location accuracy and activity recognition via evidential reasoning. This aim was supported by four key research questions/objectives. The reminder of this section will discuss these questions/objectives and how these were met throughout this thesis and the contribution to knowledge they represent.

Chapter Two presented a detailed discussion on the opportunities for contribution within the areas of supporting ADLs within a wearable computing and smart environment context. Along with providing a detailed overview of ADLs along with techniques to leverage smart environments to support ADLs and how to use technology as an enabler.

7.2.1 Research Question One

Does the use of an egocentric wearable camera offer the ability to determine the user's indoor localisation along with additional context when detecting activities in comparison to dense sensing approaches?

Chapter Two, Section 2.7.1 presented a discussion on the limitations of dense sensing apporaches for determine the user's indoor location. The main challenges that were found, the requirement for equipment to be installed throughout the environment [172, 173, 174, 175, 176, 177, 179, 180, 182, 183, 184, 186], the requirement to wear a dedicated device [5, 176, 178, 179, 183, 181, 182, 183, 184, 185, 198, 210, 204, 206, 223, 226], issues regarding multiple occupancy [172, 173], and the necessity for the occupant to interact with a sensor in order to determine location [172, 173, 179].

In Chapter Four the design and development of a solution to facilitate indoor localisation through the use of a single "always-on" egocentric camera via real time streaming was presented. This included the development of novel fiducial marker designs which could be applied to "key" objects within an environment. A review of feature point recognition algorithms also takes place. The main contribution of this Chapter was the implementation of a two-stage filtering process to reduce the number of FP detected within the video stream. The first filter calculated the number of feature points within the homography and compared these to predetermined threshold values to determine if a detection was likely to be a TP. The second stage of the filtering was a vote function, were frames were processed in batches and the object most likely to be present within these frames is determined and then stored within the system. The objects detected within each of these batches was then determined to be a TP if the number of votes exceeded a threshold value. The proposed system was tested within multiple smart environments in order to determine its feasibility to be applied to multiple environments. It was tested in both Ulster University, UK and the University of Jaèn, Spain. The results from this Chapter show that the presented system is a feasible method of determining the location of an occupant within an environment with the results from the lab at Ulster University showing a Recall, Precision, and F-Measure of 0.82, 0.96, and 0.88 respectively. The results from the lab at the University of Jaèn were also promising showing a Recall, Precision, and F-Measure of 0.66, 0.67, and 0.79 respectively.

This research question has been answered through this work both by the results from the tests within multiple environments but also by the secondary advantages this system offers. This demonstrated the ability of the system to be easily applied to multiple environment without the need for extensive equipment installation. While the occupant is still required to wear a device within this research this is somewhat mitigated by the percentage of the adult population that are required to wear corrective lenses [187]. As Google Glass can be fitted with prescription lenses the occupant does not have to wear an additional device, only substitue their current glasses for Google Glass. Occlusion, traditionally a problem of machine-vision systems [70, 188, 193, 201, 202, 204, 208, 225, 226], has been avoided through this method due to the camera being mounted on the user which provided a first-person view point, removing the issue of occlusions. An additional advantage the proposed system offers is that of continuous image capture to reduce the number of missed object interactions and missing data which was found to be a limitation of existing systems [193, 194, 202]. Finally, the issue of multiple occupancy is also negated. As the system only has to support the occupant that is wearing a device it does not need to be concerned with any additional occupants within the environment. The wearable device is streaming an egocentric view point of the occupant ensuring that any object that enters the FoV will be an object that the occupant is likely to be interacting with.

7.2.2 Research Question Two

Does the use of fiducial markers within the environment allow the easy adaption to new environments without a period of re-learning the environment?

Chapter Two, Section 2.7.2 presented a discussion on the limitations of vision based indoor localisation. A number of limitations were found, such as the need to be retrained for each environment that the system is to be deployed within [70, 192, 194, 199, 205, 206, 208, 209, 210, 211, 214, 215, 218, 220, 223, 225, 226, 227].

Chapter Four presented the novel fiducial markers that were applied to objects of interest within the environment. The markers were then applied to the smart lab environment at Ulster to perform a range of ADLs with the goal of recognising the constitute objects/locations that the occupant was interacting with. The system did not require any learning of the environment, each fiducial marker has a unique label associated with the marker (typically the object name) which is supplied at system start up. The system calculates the relevant feature points for each marker and these are stored within the system. The feature points are then compared in real-time to what is being captured by the egocentric camera using a brute-force matcher, if a suspected match is found it is passed further down the system to the filtering process detailed in Chapter 4. Chapter 5 detailed how the system could be adapted to a new environment without the need for a period of re-learning, due to the fiducial markers and their associated ID being all that was required for the system. This allowed for an easy adaption to the new environment, particularly through the use of a wearable camera as this also removes the need for equipment to be set up within the environment, such as in a dense sensing solution, along with the lack of a re-learning to the new environment.

The application of the proposed system to multiple environments without a period of re-learning demonstrates how this research question has been answered, along with secondary advantages that this solution offers. One of which is the ability to customise markers to the environment to aid in reducing any further distress or confusion for the occupants. The method of using fiducial markers to detect objects also negates the need for the system to be trained for new environments, due to the markers having an ID which associated it with the 'key' objects, should the user replace a 'key' object the system no longer needs to be retrained to learn this new object as the same marker can be applied to the new object.

7.2.3 Research Question Three

Does the use of an object-distance estimation improve the rate of detection of object interaction when compared to a non-estimation approach?

An additional limitation that became apparent throughout the course of this research was the need to reduce FP caused by navigation throughout the environment or in object rich environments. An example of this could the kitchen which could have a number of objects of interest within close proximity of each other, *e.g.*, kettle, microwave, and cupboards.

In Chapter Five the design and development of a solution to determine if an occupant/object interaction a True or False Positive was presented. The system was defined as the Intelligent System for Detecting Inhabitant-Object Interaction (ISDII). ISDII determined if an occupant/object interaction was genuine through measuring the distance that the occupant is interacting with the object and cross-referencing that against known threshold interaction distances for each object to assess if the interaction was a TP or a FP generated through general gaze activity or through navigating the environment. The presented solution offers a non-intrusive method of determining when an occupant/object interaction is genuine and not a FP. Leveraging a single wearable camera and was shown to reduce FP instances as discussed in Chapter Five. This also offers an additional advantage, such as the lack of required interaction from the occupant to record interactions as being TP/FP and that the camera is always optimised for the direction that the object of interest for the user is positioned.

7.2.4 Research Question Four

Does the application of evidential reasoning further improve the state of the art through improving the accuracy of activity recognition?

Chapter Two, Section 2.7.5 presented a discussion on the challenge of dealing with uncertainty within the data [238, 239, 240, 241, 242, 234]. In Chapter Six an implementation of DS theory was presented with the goal of determining the probability of an activity being carried out within a video stream. The goal of this Chapter was to determine if DS theory could be used to improve the accuracy of activity recognition when utilising an egocentric camera. An implementation of DS theory was applied to the datasets previously collected within the Ulster and Jaén labs to estimate the activity being carried out within the video stream. A range of ML algorithms were also tested on the datasets in order to provide a benchmark score to contextualise the results from the DS system. The DS system showed improved result in terms of number of activities correctly classified (84%) when compared to traditional ML approaches (75% for best ML approach tested) along with an improvement in recall and F-measure when compared to ML approaches – 1.00 and 0.92 respecfully for the DS system, and 0.750 and 0.778 respectfully for the best performing ML approach tested.

These results indicate that the application of evidential reasoning can further improve the accuracy of activity recognition when compared to traditional ML approaches. This is of particular importance within the context of machine-vision applications as there are a wide range of factors that can affect the detection performance, such as lighting or occlusions. The use of DS theory allows for a level of certainty to be attained for the likelihood of an activity being carried out with missed sensor events, thus allowing for a further level of accuracy to be attained than would otherwise be possible once uncertainty is introduced to the dataset.

7.2.5 Summary of Knowledge Contributions

The research carried out within this thesis has contributed to knowledge in a number of key areas. These contributions are detailed below:

The design and implementation of a real-time vision based indoor localisation system via an egocentric camera utilising fiducial markers. (Objective one)

One of the contributions offered by this work include addressing a problem previously identified with that of wearable devices such as Google Glass. That is, that their impact in ubiquitous computing and ambient intelligence systems has been partly slowed by their lack of streaming [187]. This has been addressed in Chapter 4 by the development of live streaming functionality from a wearable device, Google Glass in this case, which allows the video stream to be accessed by multiple sources using a media server. Due to the time sensitive nature of supporting occupants within their own home, a real-time system will allow a more timely and effective intervention when compared to a system which capture images on an intermittent basis or has a large time delay between the image being captured and processing being completed.

The use of an egocentric camera along with fiducial markers also aids in alleviating an issue identified in Chapter 2 such as occlusion from fixed cameras where the occupant is not within the camera's field of view due to large objects occluding the occupant or "blank" areas of the environment where the camera's field of view does not cover. While there is a risk of occlusion of the fiducial markers this is greatly reduced through the use of a first-person camera which removes the issue of covering the entire room along with large items, such as doors/fridges. This also aids in reducing occlusions generated by the occupant themselves, such as hands/head/torso occluding objects that they are interacting with. Occlusions of the manipulated object tend to be lessened as the object being interacted with is usually the centre of attention for the occupant [314]. As the object is the centre of the occupant's attention the object is usually in the centre of the image and in focus, providing a high quality image for processing [314]. Secondly, the use of a moving camera coupled with static objects reduces the issues traditionally seen with a static camera solution such as the limited field of view, which may require the installation of multiple cameras within an environment.

Due to the system operating in real-time it does not encounter the same issues as intermittent image capture system. Where vital information could be lost if the occupant interacts with an object or navigates throughout the environment. In the previously discussed works in Chapter 2 the method of image capture relied on intermittent captures, *e.g.* at set time intervals 30 frames were captured. This could cause vital information to be lost as object interactions may have taken place within the time period were the system was not capturing information. As the presented system operates in real-time every frame is being processed, therefore vital information will not be lost through intermittent image capture.

The proposed approach offers other secondary advantages that are unique to this method, such as the first person view and lack of required interaction and multiple occupancy, where each occupant that requires support need only to wear a device to obtain their unique first person viewpoint and the information on the objects they were interacting with.

The design and implementation of a method to remove the need to train for each environment. (Objective two)

The main contribution offered within objective two is the ease with which the system can be deployed within differing environments. The use of fiducial markers with an associated ID negates the need for specific training to each environment. This is due to the markers being associated with common static items that are commonly found within home environments, with the ID of the object being tied to the marker rather than any features of the object itself. This allows the system to be quickly and easily deployed within new environments in comparison to implementing traditional methods of indoor localisation. Due to the static items that the markers are applied to being common throughout the majority of homes (*e.g.* fridge, kettle, *etc.*) results in a further minimisation of the initial installation/initialisation requirements. This is due to the majority of markers sharing their ID with common household items, with only slight customisation required to any unique appliances or needs that the occupant may require. This also allows for ease of future customisation within the environment, should the occupant add additional appliances or require reduced/increased levels of support this can easily be accommodated through the addition or removal of fiducial markers.

Benchmarking ORB and Aruco in an AAL scenario along with the development of IDSII. (Objective three)

The contributions offered within objective three include the comparison of two popular off-the-shelf algorithms for feature detection in an AAL scenario. It also presents how lighting effects the performance of these two algorithms as well as that of motion blur, these are two very important factors when assessing the effectiveness of vision based aids and their feasibility in being applied to a real world situation.

ISDII is another contribution that this objective has made. This itself has two contributions within. Namely, the development of a two stage filter which allows uncertainty in real-time video based application to be reduced through exponential smoothing to reduce high frequency noise, and the second stage which involves the removal of isolated detections, such as those experienced through natural gaze activity. This used fuzzy logic to estimate the level of interaction the occupant is having with the object through distance estimation. Due to the high levels of noise that are typically present in egocentric videos it can be difficult to identify the correct object as it is possible that multiple objects can be within the occupant's FoV. This is due to some areas of the environment being densely populated with relevant objects, such as the kitchen.

A final contribution from this objective was the development of a system that does not require user interaction in order to ensure that the best image angle is being captured. This challenge was previously identified and presented in Chapter Two as a limitation of existing systems.

Implementation of DS theory to that of an egocentric camera in order to correctly identify ADLs within a real world smart environment. (Objective four)

Chapter 6 presented a methodology for applying DS theory to an egocentric camera with the goal of identifying ADLs. A comparison was also offered to traditional ML methods to establish that the use of DS theory can improve the detection of ADLs within a smart environment.

The contribution offered by this objective is an implementation of DS theory to that of an egocentric camera in order to correctly identify ADLs within a real world smart environment. This aids in alleviating the problem of unreliable sensor evidence, particularly within a machine-vision system were a high number of variables, such as light, viewing angle, *etc.*, can affect the accuracy of an object recognition system. The advantage of detecting ADLs, even with missing sensor values, offers increased reliability and safety with the domain of ambient assisted living and further moving towards a vision of "aging in place". Furthermore new activities can be added to the system without the need for existing data on the activities present due to the activities being composed of a composite of unique objects, this allowing a new activity to be added by merely including the composite objects in the activity template.

7.3 Limitations

Over the course of this thesis a number of limitations have been encountered, these include:

7.3.1 Marker Design

The novel design of the fiducial markers within this thesis does offer the benefit of being able to customise the marker not just to the environment, but also to the unique needs of the occupant. However, it became apparent that there may be issues in regards to scalability, this is due in part to a certain level of complexity being required in order to identify a marker within a scene. While this was not an issue during this research if the system was to be scaled out to multiple environments it could result in markers requiring to have overly complex designs in order to easily distinguish them for similarly designed markers due to the ORB algorithm being based on the Harris corner recognition algorithm which extracts corners to infer the features of an image. An additional limitation is that of the requirement for the marker to be placed correctly on the corresponding object to facilitate the accurate detection of the marker. This is somewhat mitigated through the markers requiring to be placed in the centre of the occupant's FoV, however, there is still room for human error when placing the markers. Alongside the potential interpretation of the centre of the occupant's FoV, particularly if being set up by a carer or family member.

7.3.2 Dataset Size and Users

A key limitation of this research has been the dataset size used throughout. This was due to multiple factors. A key factor was the use of novel fiducial markers as no existing egocentric dataset were available, therefore a fresh dataset had to be collected throughout this research. Additionally collecting egocentric data of ADLs is a time consuming process which when coupled with hardware limitations discussed in Section 7.3.3 further hinders the collection of a large dataset. The small nature of the dataset had the most impact on Chapter 6 in which ML approaches were applied to the collected datasets in order to determine the activity being carried out which resulted in a test/train split for validation to avoid any bias in the results or overfitting to the data. Further collection of a larger dataset would allow for additional testing to establish if the results can be further generalised. This would also allow the opportunity to test the system in additional environments thus allowing the hypothesis of the *"use of fiducial markers facilitates the ease of adaption to new environments"*, while allowing testing against a further range of environmental conditions. A further limitation within the datasets gathered in the lack of data generated by target users. The datasets generated in Chapter Three were generated by a researcher which was not a member of the target cohort of older users which should be a key focus of future work. While data augmentation was considered it was felt that the limitations of such a solution would outweigh the benefits. There was a concern that the data quality could be affected due to the generation of unrealistic or irrelevant data. Additionally, as data augmentation can only generate variations of the existing data would result in limited diversity within the dataset despite the increased dataset size. As data augmentation cannot create new, original data no new features/information which was not in the original dataset would be generated. This was of particular interest within Chapter Six where defects/inconsistencies in the data were of interest to explore how DS theory could manage missing/inconsistent data.

7.3.3 Hardware Limitations

During the course of this research key hardware limitations became apparent. Firstly, the act of streaming live video over a continuous time period led to issues with heat management on the Google Glass device. In order to compensate for the increased heat the Google Glass device reduced the clock speed of the CPU. The first step Google Glass takes is to reduce the CPU speed from 1Ghz to 600Mhz. Should this not be successful at mitigating the increased temperatures Google

Glass can then further reduce the CPU speed from 600Mhz to 300Mhz. However, it should be noted that the device did not become too hot to be a danger or uncomfortable for the user to remain wearing/using, though this did have the consequence of introducing a slight variable lag into the video stream when the clock speed was reduced to 300Mhz of approximately three seconds or less. A secondary limitation that was discovered was that of battery capacity due to the energy intensive requirements of both recording a constant video stream while simultaneously streaming to a server.

7.3.4 Time Frame

One final limitation to this research is the period of time it has taken to bring this Thesis to completion, in particular the domain of Computer Science has a very high rate of progress with regards to the tools and technologies that are available. With regards to the machine-vision aspect of this research techniques such as Convolutional Neural Networks (CNN) [330] and more recently the development of Vision Transformers (ViT) [331] offer intriguing aspects into the future of machine-vision applications. However, the domain of AAL still has significant research challenges with smart environments still not widely available, it is hoped this research will aid in informing this future research.

7.4 Future Work

With the opportunity for reflection on the work conducted, and its outcomes, throughout this thesis a number of areas of future work have been identified.

7.4.1 Further Data Collection and Deployment

As discussed in Chapter 6 both ML and DS theory were applied to the collected datasets from the Ulster and Jaén smart environments, however, these datasets were relatively small within the context of data science containing a total of 64 events. Further data collection would allow for further generalisation of the ML models along with offering further opportunity to evaluate how effectively DS theory can be applied to activity recognition within the domain of AAL. Collection of further datasets would also permit the opportunity to further assess the ease of adaption to new environments of varying complexity. The collection of additional data would also allow the system to be implemented in a free-living environment outside of a laboratory setting, this would allow data to be collected by end users in their home environment and would allow further evaluation on the effectiveness of the ISDII component to detect genuine occupant/object interactions. A key component of further data collection should be the inclusion of users from the target cohort of the older segement of society. This will allow the collection of data from a real world scenario using target users that will offer valuable insights into the usability and acceptability of the system by the target cohort. This would also permit further testing on the ability of DS theory to deal with uncertainty within the data alongside allowing for end-user feedback on the system and any recommendations to aid in the widespread adoption of such technology.

7.4.2 Activity Support

This thesis has explored methods of detecting the occupants' location within an environment through the use of a wearable camera and the detection of fiducial markers via machine-vision techniques and if this system could be adapted to any environment without the need for training. This was followed by an investigation into the development of the ISDII system in order to determine if an object interaction was genuine or caused by navigation throughout the environment or due to a high concentration of objects of interest within the FoV. Finally a study was carried out to determine if DS theory could be implemented to correctly identify ADLs being carried out from an egocentric viewpoint within a smart environment taking into account uncertainty introduced to the data.

Further work in this area could involve feeding information back to the occupants to assist the occupant in their daily routines. In particular context-aware reminders could provide a valuable service to older users, or those who may have early stages of cognitive decline along with aiding the occupant in completing their current task or prompting them to begin a task, such as making a meal. As Google Glass contains an integrated bone conduction speaker along with a 640x360 display this opens up a range of possible reminders which can be delivered to the occupant. This is of particular interest if they are suffering from cognitive decline, such as early onset dementia, as it allows the reminders to be delivered in a format that would offer the least stress to the occupant. These reminders could take the format of video/audio recording of family members who are known to the occupant to make them feel more at ease with assistive technology.

7.4.3 Digital Twins

One interesting avenue of investigation is that of leveraging a digital twin to aid in supporting occupants within their own home. A digtial twin could allow for a accurate representation of the environment including the position of the fiducial markers along with the position of objects within the environment. This digital recreation of the environment can facilitate the testing of new activity models. Alongside investigating issues such as multiple occupancy in an evironment that allows the mitigation of risks when compared to testing in a real world environment. Additionally, this can be used to generate further training data that simulates a real world environment in a range of contextual situations.

The use of a digital twin would also allow for additional performance evaluations to take place through introducing specific challenges to a scenario. For example, the lighting conditions could be varied, occlusions could be introduced, along with varying the location of the marker placement to assess how the system performance is affected. This would also allow the calibration of the system in real time by comparing the simulated environment with the real world environment. The system would then be able to adjust it's parameters to match the current real world conditions by adjusting camera settings. Such as increasing exposure to aid in low light environments. This can allow performance metrics to be established to aid improving the accuracy of indoor localisation and activity recognition within assistive technologies.

7.4.4 Summary

In conclusion, future work for this research would involve the integration of the areas that have been identified within this section – further data collection, activity support, and digital twins. Further data collection would allow a more diverse dataset to be gathered, both in terms of varying environmental conditions and activities carried out. Further data would also allow an investigation into utilising vision transformers as a means of improving the detection accuracy of the system. The integration of a digital twin would allow for additional testing and allow for mulitple interations of the model to be tested under varying conditions. The inclusion of a digital twin would also allow the real time synchronisation between the digital twin and the real world device, this would allow modifications to be made in real time to aid in improving performance through the adjustment of system settings. Finally, the built in screen to smart glasses could be leveraged to aid giving further support to the occupant to aid them in completing tasks via visual and audio prompts.

7.5 Conclusion

This Thesis explored an investigation into whether machine-vision baed approaches could be leveraged to support indoor localisation within the domain of AAL. This Thesis then went on to investigate could the application of evidential reasoning via DS theory could improve the detection of an occupant's activity within a smart environment. Study One carried out within this research was an investigation to determine if an egocentric camera could be utilised to determine an occupant's location based on the objects within their FoV as detailed within Chapter Four. Study Two sought to determine if it was possible to develop a tool which would allow occupant/object interactions to be determined to be genuine or accidental. To this end the ISDII tool was developed which used distance estimation to make a determination if the object was likely to be a genuine interaction based off expert defined distances that objects were typically interacted upon with the end goal of reducing the number of FPs detected within the video stream. Study Three aimed to detect the activity that the occupant was carrying out, along with implementing DS theory in order to aid with dealing with uncertainty within the data which is normally present in a real world scenario due to technical faults or user error.

Throughout the research conducted within this Thesis a number of contributions to knowledge have been identified. These have been a direct result of the overall research aim of this Thesis, namely to "*investigate the use of machine-vision based approaches to support those at home who may traditionally require assistance to carry out their activities of daily living through the use of improved location accuracy and activity recognition via evidential reasoning*". Contributions from this Thesis have been discussed in Chapters 4, 5, and 6 and have been outlined in Section 7.2.5. The studies detailed in Chapters 4, 5, and 6 have allowed the research objectives of this Thesis to be achieved, Section 7.2 discusses these research objects in more detail and how they were achieved through the course of this research.

This Thesis has also highlighted some limitations within this research which have been discussed in Section 7.3 of this Chapter. The limitations included the novel marker design, which while offering customisability to suit an environment or the occupant's needs can result in scalability issues depending on the size of environment/number of objects that are required to be supported by the system. The dataset size was also a limitation within this research, particularly when implementing the ML component within Chapter Six. Further expansion of the dataset would also allow for additional testing in differing environments to further demonstrate the system's ability to be applied to new environments with no need for training. Additionally hardware limitations were also discovered, with Google Glass requiring to under-clock the processor in order to reduce heat output of the device. However, with the latest generation of devices such as Google Glass Enterprise [168] and Vuzix Blade [31] these hardware limitations will be reduced through the progress made within the IoT domain since the inception of Google Glass Explorer.

It is hoped that this Thesis will aid in the development of future applications within the domain of AAL to support aging-in-place and to further contribute to the vision of Mark Weiser of ubiquitous computing offering on-demand support seemlessly within our lives.

References

[1] American Occupation Therapy Association, "Occupational Therapy Practice Framework: Domain and Process," *The American Journal of Occupational Therapy*, vol. 56, no. 6, pp. 609–639, 2002.

[2] H.-F. Mao, L.-H. Chang, G. Yao, W.-Y. Chen, and W.-N. W. Huang, "Indicators of perceived useful dementia care assistive technology: Caregivers' perspectives.," *Geriatrics & Gerontology International*, vol. 8, no. 32, pp. 1–9, 2014.

[3] N. Camp, M. Lewis, K. Hunter, J. Johnston, M. Zecca, A. Di Nuovo, and D. Magistro, "Technology used to recognize activities of daily living in community-dwelling older adults," *International Journal of Environmental Research and Public Health*, vol. 18, no. 1, pp. 1–18, 2021.

[4] A. Sanchez-Comas, K. Synnes, and J. Hallberg, "Hardware for recognition of human activities: A review of smart home and AAL related technologies," *Sensors (Switzerland)*, vol. 20, no. 15, pp. 1–26, 2020.

[5] Ó. Belmonte-Fernández, A. Puertas-Cabedo, J. Torres-Sospedra, R. Montoliu-Colás, and S. Trilles-Oliver, "An indoor positioning system based on wearables for ambient-assisted living," *Sensors (Switzerland)*, vol. 17, no. 1, pp. 1–22, 2017.

[6] K. Ngamakeur, S. Yongchareon, J. Yu, and S. U. Rehman, "A Survey on Device-free Indoor Localization and Tracking in the Multi-resident Environment," *ACM Computing Surveys*, vol. 53, no. 4, 2020.

[7] W. C. S. S. Sim, G. S. Machado, A. M. A. Sales, and M. M. D. Lucena, "A Review of Technologies and Techniques for Indoor," *Sensors*, vol. 20, no. 14, p. 3935, 2020.

[8] F. Zafari, A. Gkelias, and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," *arXiv*, vol. 21, no. 3, pp. 2568–2599, 2017.

[9] K. Ha, Z. Chen, W. Hu, W. Richter, P. Pillai, and M. Satyanarayanan, "Towards wearable cognitive assistance," in *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*, pp. 68–81, ACM, 2014.

[10] S. A. K. Tareen and Z. Saleem, "A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK," in *2018 International Conference on Computing, Mathematics and Engineering Technologies: Invent, Innovate and Integrate for Socioeconomic Development, iCoMET 2018 - Proceedings*, vol. 2018-Janua, pp. 1–10, IEEE, 2018.

[11] A. J. B. Brush, B. Lee, R. Mahajan, S. Agarwal, S. Saroiu, and C. Dixon, "Home Automation in the Wild: Challenges and Opportunities," in *CHI Conference on Human Factors in Computing Systems*, pp. 2115–2124, 2011.

[12] X. Hong, C. Nugent, M. Mulvenna, S. McClean, B. Scotney, and S. Devlin, "Evidential fusion of sensor data for activity recognition in smart homes," *Pervasive and Mobile Computing*, vol. 5, no. 3, pp. 236–252, 2009.

[13] K. Christensen, G. Doblhammer, R. Rau, and J. W. Vaupel, "Ageing populations: the challenges ahead.," *Lancet*, vol. 374, pp. 1196–208, oct 2009.

[14] D. J. Cook and S. K. Das, "How smart are our environments? An updated look at the state of the art," *Pervasive and Mobile Computing*, vol. 3, no. 2, pp. 53–73, 2007.

[15] M. Weiser, "The Computer for the 21st Century," *Scientific American*, vol. 265, no. 3, pp. 94—-104, 1991.

[16] K. Courtney, G. Demiris, M. Rantz, and M. Skubic, "Needing smart home technologies : the perspectives of older adults in continuing care retirement communities," *Informatics in Primary Care*, vol. 16, no. 3, pp. 195–201, 2008.

[17] H. Sun, V. D. Florio, N. Gui, and C. Blondia, "Promises and Challenges of Ambient Assisted Living Systems," in *2009 Sixth International Conference on Information Technology: New Generations*, pp. 1201–1207, IEEE, 2020.

[18] V. Nathan, S. Paul, T. Prioleau, L. Niu, B. J. Mortazavi, S. A. Cambone, A. Veeraraghavan, and A. Sabharwal, "A Survey on Smart Homes for Aging in Place: Toward Solutions to the Specific Needs of the Elderly," *IEEE Signal Processing Magazine*, vol. 35, no. August, pp. 111–119, 2018.

[19] A. K. Dey and G. D. Abowd, "Towards a Better Understanding of Context and Context-Awareness," in *What, Who, Where, When, Why and How of Context-Awareness*, (Rotterdam), pp. 304–307, 2000.

[20] A. Sandy, "Healthwear: medical technology becomes wearable," *Computer*, vol. 37, no. 5, pp. 42–49, 2004.

[21] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation.," *Journal of neuroengineering and rehabilitation*, vol. 9, p. 21, jan 2012.

[22] J. Zhu, K. Shi, C. Yang, Y. Niu, Y. Zeng, N. Zhang, T. Liu, and C. H. Chu, "Ethical issues of smart home-based elderly care: A scoping review," *Journal of Nursing Management*, vol. 30, no. 8, pp. 3686–3699, 2022.

[23] C. Berridge, T. F. Wetle, and B. J. Bowers, "Why Older Adults and Their Children Disagree about In-Home Surveillance Technology, Sensors, and Tracking," *Gerontologist*, vol. 60, no. 5, pp. 926–934, 2020.

[24] Y. G. Park, S. Lee, and J. U. Park, "Recent progress in wireless sensors for wearable electronics," *Sensors (Switzerland)*, vol. 19, no. 20, pp. 1–34, 2019.

[25] M. Haghi, K. Thurow, and R. Stoll, "Wearable devices in medical internet of things: Scientific research and commercially available devices," *Healthcare Informatics Research*, vol. 23, no. 1, pp. 4–15, 2017.

[26] V. Vijayan, J. Connolly, J. Condell, N. McKelvey, and P. Gardiner, "Review of wearable devices and data collection considerations for connected health," *Sensors*, vol. 21, no. 16, pp. 1–31, 2021.

[27] T. Basaklar, Y. Tuncel, S. An, and U. Ogras, "Wearable Devices and Low-Power Design for Smart Health Applications: Challenges and Opportunities," *Proceedings of the International Symposium on Low Power Electronics and Design*, vol. 2021-July, p. 1, 2021.

[28] L. Lu, J. Zhang, Y. Xie, F. Gao, S. Xu, X. Wu, and Z. Ye, "Wearable health devices in health care: Narrative systematic review," *JMIR mHealth and uHealth*, vol. 8, no. 11, 2020.

[29] Google, "Google Glass," 2017, http://www.x.company/glass/, (accessed Nov. 22, 2017).

[30] J. Meyer, A. Frank, T. Schlebusch, and E. Kasneci, "A CNN-based Human Activity Recognition System Combining a Laser Feedback Interferometry Eye Movement Sensor and an IMU for Context-aware Smart Glasses," in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, 2021.

[31] Vuzix, "Vuzix Blade," 2022, https://vuzix.uk/, (accessed Oct. 2, 2023).

[32] Ray-Ban, "Ray-Ban Stories," 2022, https://www.ray-ban.com/, (accessed Dec. 7, 2022).

[33] Bose, "Bose Frames," 2022, https://www.bose.com/, (accessed Dec. 7, 2022).

[34] SnapChat, "Snap Spectacles 3," 2022, https://www.spectacles.com/, (accessed Dec. 7, 2022).

[35] Amazon, "Amazon Echo Frames," 2022, https://www.amazon.com/, (accessed Dec. 12, 2022).

[36] Google, "Glass Enterprise Edition 2," 2021, https://www.google.com/glass/tech-specs/, (accessed Oct. 2, 2023).

[37] ReportLinker, "Smart Glasses Market – Global Industry Size, Share, Trends, Opportunity and Forecast, 2018-2028," 2023, (accessed Jul. 25, 2023).

[38] Xiaomipedia, "Xiaomi Glasses Camera," 2022, https://xiaomipedia.com/, (accessed Jul. 25, 2023).

[39] EE, "Nreal Air," 2022, https://ee.co.uk/, (accessed Jul. 25, 2023).

[40] R. J. Shei, I. G. Holder, A. S. Oumsang, B. A. Paris, and H. L. Paris, "Wearable activity trackers–advanced technology or advanced marketing?," *European Journal of Applied Physiology*, vol. 122, no. 9, pp. 1975–1990, 2022.

[41] P. Brown, J. Bovey, and C. Xian, "Context-aware applications: from the laboratory to the marketplace," *Personal Communications, IEEE*, vol. 4, no. 5, pp. 58–64, 1997.

[42] A. Almeida, R. Mulero, P. Rametta, V. Urošević, M. Andrić, and L. Patrono, "A critical analysis of an IoT—aware AAL system for elderly monitoring," *Future Generation Computer Systems*, vol. 97, pp. 598–619, 2019.

[43] Y. Rahal, H. Pigot, and P. Mabilleau, "Location estimation in a smart home: system implementation and evaluation using experimental data.," *International journal of telemedicine and applications*, vol. 2008, p. 142803, jan 2008.

[44] H.-Y. Noh, J.-H. Lee, S.-W. Oh, K.-S. Hwang, and S.-B. Cho, "Exploiting indoor location and mobile information for context-awareness service," *Information Processing & Management*, vol. 48, pp. 1–12, jan 2012.

[45] J. Kunhoth, A. G. Karkar, S. Al-Maadeed, and A. Al-Ali, "Indoor positioning and wayfinding systems: a survey," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, 2020.

[46] H. Obeidat, W. Shuaieb, O. Obeidat, and R. Abd-Alhameed, *A Review of Indoor Localization Techniques and Wireless Technologies*, vol. 119. Springer US, 2021.

[47] J. Hightower and G. Borriello, "Systems for Ubiquitous Computing," *Computer*, vol. 34, no. 8, pp. 57–66, 2001.

[48] J. Gómez-Romero, M. a. Serrano, M. a. Patricio, J. García, and J. M. Molina, "Context-based scene recognition from visual data in smart homes: an Information Fusion approach," *Personal and Ubiquitous Computing*, vol. 16, pp. 835–857, sep 2011.

[49] A. Morar, A. Moldoveanu, I. Mocanu, F. Moldoveanu, I. E. Radoi, V. Asavei, A. Gradinaru, and A. Butean, "A comprehensive survey of indoor localization methods based on computer vision," *Sensors (Switzerland)*, vol. 20, no. 9, pp. 1–36, 2020.

[50] A. Diallo, Z. Lu, and X. Zhao, "Wireless Indoor Localization Using Passive RFID Tags," *Procedia Computer Science*, vol. 155, no. 2018, pp. 210–217, 2019.

[51] C. H. Cheng, M. F. Wu, S. H. Guo, and K. S. Chen, "Realization and performance assessment of a force-sensing based smart floor panel for indoor localization applications," *Sensors and Actuators, A: Physical*, vol. 285, pp. 468–481, 2019.

[52] L. Bai, F. Ciravegna, R. Bond, and M. Mulvenna, "A Low Cost Indoor Positioning System Using Bluetooth Low Energy," *IEEE Access*, vol. 8, pp. 136858–136871, 2020.

[53] Z. Chen, H. Zou, J. Yang, H. Jiang, and L. Xie, "WiFi Fingerprinting Indoor Localization Using Local Feature-Based Deep LSTM," *IEEE Systems Journal*, vol. PP, pp. 1–10, 2019.

[54] T. Kim Geok, K. Zar Aung, M. Sandar Aung, M. Thu Soe, A. Abdaziz, C. Pao Liew, F. Hossain, C. P. Tso, and W. H. Yong, "Review of Indoor Positioning: Radio Wave Technology," *Applied Sciences*, vol. 11, no. 1, p. 279, 2020.

[55] Y. Yu, R. Chen, L. Chen, S. Xu, W. Li, Y. Wu, and H. Zhou, "Precise 3-D Indoor Localization Based on Wi-Fi FTM and Built-In Sensors," *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 11753–11765, 2020.

[56] A. K. Dey, "Context-Aware Computing," in *Ubiquitous Comuting Fundamentals* (J. Krumm, ed.), ch. 8, pp. 321–352, Chapman and Hall/CRC, first ed., 2009.

[57] T. Le Dinh, T. T. Pham Thi, C. Pham-Nguyen, and L. N. H. Nam, "A knowledge-based model for context-aware smart service systems," *Journal of Information and Telecommunication*, vol. 6, no. 2, pp. 141–162, 2022.

[58] S. Spinsante, A. Poli, J. Mongay Batalla, P. Krawiec, C. Dobre, L. Băjenaru, C. X. Mavromoustakis, C. S. Costantinou, G. Molan, A. M. Herghelegiu, G. I. Prada, R. Drăghici, and H. González–Vélez, "Clinically-validated technologies for assisted living: The vINCI project," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 3, pp. 2095–2116, 2023.

[59] M. Baldauf, "A survey on context-aware systems," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 2, no. 4, pp. 263–277, 2007.

[60] V. R. Chifu, C. B. Pop, D. Demjen, R. Socaci, D. Todea, M. Antal, T. Cioara, I. Anghel, and C. Antal, "Identifying and Monitoring the Daily Routine of Seniors Living at Home," *Sensors*, vol. 22, no. 3, 2022.

[61] A. Patel and J. Shah, "Real-time human behaviour monitoring using hybrid ambient assisted living framework," *Journal of Reliable Intelligent Environments*, vol. 6, no. 2, pp. 95–106, 2020.

[62] L. Correia, D. Fuentes, J. Ribeiro, N. Costa, A. Reis, C. Rabadão, J. Barroso, and A. Pereira, "Usability of smartbands by the elderly population in the context of ambient assisted living applications," *Electronics (Switzerland)*, vol. 10, no. 14, pp. 1–18, 2021.

[63] N. Cha, A. Kim, C. Y. Park, S. Kang, M. Park, J. G. Lee, S. Lee, and U. Lee, "Hello There! Is Now a Good Time to Talk?:Opportune Moments for Proactive Interactions with Smart Speakers," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, 2020.

[64] G. Cicirelli, R. Marani, A. Petitti, A. Milella, and T. D'orazio, "Ambient assisted living: A review of technologies, methodologies and future perspectives for healthy aging of population," *Sensors*, vol. 21, no. 10, 2021.

[65] L. Chen and I. Khalil, "Activity Recognition: Approaches, Practices and Trends," in *Activity Recognition in Pervasive Intelligent Environments*, vol. 4, pp. 1–31, 2011.

[66] M. Wallace and M. Shelkey, "Katz Index of Independence in Activities of Daily Living (ADL).," *Urologic nursing : official journal of the American Urological Association Allied*, vol. 27, no. 1, pp. 93–94, 2007.

[67] L. Minh Dang, K. Min, H. Wang, M. Jalil Piran, C. Hee Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognition*, vol. 108, 2020.

[68] T. Pavlidis, "The challenge of general machine vision," *Signal, Image and Video Processing*, vol. 8, pp. 191–195, oct 2013.

[69] A. Rosenfeld, "Introduction to Machine Vision," *IEEE Control Systems Magazine*, vol. 5, no. 3, pp. 14–17, 1985.

[70] M. Buzzelli, A. Albé, and G. Ciocca, "A vision-based system for monitoring elderly people at home," *Applied Sciences (Switzerland)*, vol. 10, no. 1, pp. 1–25, 2020.

[71] Y. K. Choi, H. J. Thompson, and G. Demiris, "Use of an internet-of-things smart home system for healthy aging in older adults in residential settings: Pilot feasibility study," *JMIR Aging*, vol. 3, no. 2, 2020.

[72] T. T. Zin, Y. Htet, Y. Akagi, H. Tamura, K. Kondo, S. Araki, and E. Chosa, "Real-time action recognition system for elderly people using stereo depth camera," *Sensors*, vol. 21, no. 17, 2021.

[73] S. D. Achirei, M. C. Heghea, R. G. Lupu, and V. I. Manta, "Human Activity Recognition for Assisted Living Based on Scene Understanding," *Applied Sciences (Switzerland)*, vol. 12, no. 21, 2022.

[74] B. Přibyl, A. Chalmers, and P. Zemčík, "Feature Point Detection under Extreme Lighting Conditions," in *Conference on Computer Graphics*, no. May, pp. 156–163, 2012.

[75] J. Li and N. M. Allinson, "A comprehensive review of current local features for computer vision," *Neurocomputing*, vol. 71, no. 10-12, pp. 1771–1787, 2008.

[76] A. Howedi, A. Lotfi, and A. Pourabdollah, "Exploring entropy measurements to identify multi-occupancy in activities of daily living," *Entropy*, vol. 21, no. 4, 2019.

[77] J. Le Kernec, F. Fioranelli, C. Ding, H. Zhao, L. Sun, H. Hong, O. Romain, and J. Lorandel, "Radar Sensing in Assisted Living: An Overview," *IEEE MTT-S 2019 International Microwave Biomedical Conference, IMBioC 2019 - Proceedings*, no. 1, 2019.

[78] D. Kin, K. Wang, S. Sclaroff, and K. Saenko, "A Broad Study of Pre-training for Domain Generalization and Adaptation," in *European Conference on Computer Vision*, pp. 621 – 638, Springer, 2022.

[79] K. Foley, J. Gentry, T. Myers, L. Rankin, and K. Shepard, "Activities of Daily Living (ADL) Strategies to Age in Place," *The American Journal of Occupational Therapy*, vol. 73, no. 4, 2019.

[80] L. C. Kobayashi, J. Wardle, and C. von Wagner, "Internet use, social engagement and health literacy decline during ageing in a longitudinal cohort of older English adults," *Journal of Epidemiology & Community Health*, vol. 69, no. 3, pp. 278–283, 2014.

[81] S. Chowdhury and P. pratim Chakraborty, "Universal health coverage - There is more to it than meets the eye," *Journal of Family Medicine and Primary Care*, vol. 6, no. 2, pp. 169–170, 2017.

[82] L. S. Noelker and R. Browdie, "Sidney Katz, MD: A new paradigm for chronic illness and long-term care," *Gerontologist*, vol. 54, no. 1, pp. 13–20, 2014.

[83] B. M. Martínez-Hernández, O. Rosas-Carrasco, M. López-Teros, A. González-Rocha, P. Muñoz-Aguirre, R. Palazuelos-González, A. Ortíz-Rodríguez, A. Luna-López, and E. Denova-Gutiérrez, "Association

between physical activity and physical and functional performance in non-institutionalized Mexican older adults: a cohort study," *BMC Geriatrics*, vol. 22, no. 1, pp. 1–11, 2022.

[84] C. M. Giebel, C. Sutcliffe, M. Stolt, S. Karlsson, A. Renom-Guiteras, M. Soto, H. Verbeek, A. Zabalegui, and D. Challis, "Deterioration of basic activities of daily living and their impact on quality of life across different cognitive stages of dementia: a European study," *International Psychogeriatrics*, vol. 26, no. 08, pp. 1283–1293, 2014.

[85] S. Beltz, S. Gloystein, T. Litschko, S. Laag, and N. van den Berg, "Multivariate analysis of independent determinants of ADL/IADL and quality of life in the elderly," *BMC Geriatrics*, vol. 22, no. 1, pp. 1–16, 2022.

[86] B. J. Gurland and M. S. Maurer, "Life and Works of Sidney Katz, MD: A Life Marked by Fundamental Discovery," *Journal of American Medical Directors Association*, vol. 13, no. 9, pp. 764–765, 2012.

[87] R. S. Bucks and J. Haworth, "Bristol Activities of Daily Living Scale: a critical evaluation," *Expert Review of Neurotherapeutics*, vol. 2, no. 5, pp. 669–676, 2002.

[88] L. Liu, C. Daum, N. Neubauer, A. Miguel-Cruz, and A. Rios-Rincon, "Technology to Facilitate Independence in Self Care -ADL and IADL," in *Autonomy and Independence. Synthesis Lectures on Technology and Health*, Springer, 2022.

[89] J. Berrett, A. de Kruiff, S. Pedell, and A. Reilly, "Augmented assistive technology: the importance of tailoring technology solutions for people living with dementia at home," *International Journal of Human Computer Studies*, vol. 165, no. March 2021, p. 102852, 2022.

[90] M. Assim and A. Al-Omary, "Design and Implementation of Smart Home using WSN and IoT Technologies," in *2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies, 3ICT 2020*, pp. 5–10, 2020.

[91] A. Banafa, "Edge Computing Paradigm," *Quantum Computing and Other Transformative Technologies*, pp. 81–85, 2023.

[92] I. Machorro-Cano, G. Alor-Hernández, M. A. Paredes-Valverde, L. Rodríguez-Mazahua, J. L. Sánchez-Cervantes, and J. O. Olmedo-Aguirre, "HEMS-IoT: A big data and machine learning-based smart home system for energy saving," *Energies*, vol. 13, no. 5, 2020.

[93] C. Rodríguez-Gallego, F. Díez-Muñoz, M. L. Martín-Ruiz, A. M. Gabaldón, M. Dolón-Poza, and I. Pau, "A collaborative semantic framework based on activities for the development of applications in Smart Home living labs," *Future Generation Computer Systems*, vol. 140, pp. 450–465, 2023.

[94] M. Bhatia, S. Kaur, and S. K. Sood, "IoT-inspired smart home based urine infection prediction," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 5, pp. 5249–5263, 2023.

[95] N. Manikandan, S. Nitisha Tadiboina, M. Shais Khan, V. Prakash, R. Singh, and K. Kumar Gupta, "Automation of Smart Home for the Wellbeing of Elders Using Empirical Big Data Analysis," in *International Conference on Advance Computing and Innovative Technologies in Engineering*, pp. 1164–1168, 2023.

[96] C. Wilson, T. Hargreaves, and R. Hauxwell-Baldwin, "Smart homes and their users: a systematic analysis and key challenges," *Personal and Ubiquitous Computing*, vol. 19, no. 2, pp. 463–476, 2014.

[97] G. Gibson, L. Newton, G. Pritchard, T. Finch, K. Brittain, and L. Robinson, "The provision of assistive technology products and services for people with dementia in the United Kingdom.," *Dementia*, pp. 1–21, 2014.

[98] F. Cardinaux, B. Deepayan, A. Charith, M. S. Hawley, S. Mark, D. Bhowmik, and C. Abhayaratne, "Video Based Technology for Ambient Assisted Living : A review of the literature," *Journal of Ambient Intelligence and Smart Environments (JAISE)*, vol. 1364, no. 3, pp. 253–269, 2011.

[99] X. Hong and C. Nugent, "Implementing evidential activity recognition in sensorised homes.," *Technology and Health Care*, vol. 19, no. 1, pp. 37–52, 2011.

[100] A. Cesta, G. Cortellessa, F. Fracasso, A. Orlandini, and M. Turno, "User needs and preferences on AAL systems that support older adults and their carers," *Journal of Ambient Intelligence and Smart Environments*, vol. 10, no. 1, pp. 49–70, 2018.

[101] S. Cahill, J. Macijauskiene, A.-M. Nygard, J.-P. Faulkner, and I. Hagen, "Technology in dementia care," *Journal of Dementia Care*, vol. 16, no. 5, p. 14, 2008.

[102] S. Mehrabian, J. Extra, Y. Wu, M. Pino, L. Traykov, and A. Rigaud, "The perceptions of cognitively impaired patients and their caregivers of a home telecare system," *Medical Devices: Evidence and Research*, vol. 8, pp. 21–29, 2015.

[103] L. Nchena, "Smart House Assistive Technologies for Senior Citizens," *2022 12th International Conference on Advanced Computer Information Technologies, ACIT 2022*, pp. 448–453, 2022.

[104] I. Cleland, T. Patterson, C. Nugent, F. Cruciani, N. Black, and C. Paggetti, "Doorstep Security ; using a technology based solution for the prevention of doorstep crime," in *International Workconference on Ambient Assisted Living*, pp. 2–5, Springer International Publishing, 2014.

[105] S. Ezatzadeh, M. R. Keyvanpour, and S. V. Shojaedini, "A human fall detection framework based on multi-camera fusion," *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 34, no. 6, pp. 905–924, 2022.

[106] J. Yuan, K. K. Tan, T. H. Lee, G. Choon, and H. Koh, "Power-Efficient Interrupt-Driven Algorithms for Fall Detection and Classification of Activities of Daily Living," *Sensors*, vol. 15, no. 3, pp. 1377–1387, 2015.

[107] T.-j. Chen, H.-j. Li, and J. Li, "The effects of reminiscence therapy on depressive symptoms of Chinese elderly: study protocol of a randomized controlled trial.," *BMC psychiatry*, vol. 12, no. 1, p. 189, 2012.

[108] S. Rao and A. Khan, "Effects of emotion therapies on people with dementia : A PRISMA-guided systematic review," in *2022 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 1760–1766, IEEE, 2022.

[109] P. P. K. Kor, L. L. Parial, C. T. K. Yu, J. Y. W. Liu, D. P. M. Liu, and J. M. K. Hon, "Effects of a family caregiver-delivered multi-sensory cognitive stimulation intervention for older people with dementia during COVID-19: A randomized controlled trial," *The Gerontologist*, no. May, pp. 1–12, 2023.

[110] M. K. Scullin, W. E. Jones, R. Phenis, S. Beevers, S. Rosen, K. Dinh, A. Kiselica, F. J. Keefe, and J. F. Benge, "Using smartphone technology

to improve prospective memory functioning: A randomized controlled trial," *Journal of the American Geriatrics Society*, vol. 70, no. 2, pp. 459–469, 2022.

[111] K. Hackett, S. Lehman, R. Divers, M. Ambrogi, L. Gomes, C. C. Tan, and T. Giovannetti, "Remind Me To Remember: A pilot study of a novel smartphone reminder application for older adults with dementia and mild cognitive impairment," *Neuropsychological Rehabilitation*, pp. 1–29, 2020.

[112] S. Czarnuch, S. R. Cohen, V. Parameswaran, and A. Mihailidis, "A real-world deployment of the COACH prompting system," *Journal of Ambient Intelligence and Smart Environments*, vol. 5, no. 5, pp. 463–478, 2013.

[113] B. G. Acampora, S. M. Ieee, D. J. Cook, F. Ieee, P. Rashidi, M. Ieee, and A. V. Vasilakos, "A Survey on Ambient Intelligence in Healthcare," *Proceedings of the IEEE*, vol. 101, no. 12, pp. 1–25, 2013.

[114] D. J. Cook, J. C. Augusto, and V. R. Jakkula, "Ambient intelligence: Technologies, applications, and opportunities," *Pervasive and Mobile Computing*, vol. 5, no. 4, pp. 277–298, 2009.

[115] F. Sadri, "Ambient intelligence," *ACM Computing Surveys*, vol. 43, no. 4, pp. 1–66, 2011.

[116] V. Kishor Sonti and G. Sundari, "Ambient Intelligence: An Emerging Innovation of Sensing and Service Systems," in *Applied Edge AI*, p. 22, 2022.

[117] P. Lago, C. Jimenez-Guarin, and C. Roncancio, "A Case Study on the Analysis of Behavior Patterns and Pattern Changes in Smart Environments," in *International Workconference on Ambient Assisted Living*, pp. 296–303, 2014.

[118] Y. Wang, X. Wang, D. Arifoglu, C. Lu, A. Bouchachia, Y. Geng, and G. Zheng, "A Survey on Ambient Sensor-Based Abnormal Behaviour Detection for Elderly People in Healthcare," *Electronics (Switzerland)*, vol. 12, no. 7, pp. 1–25, 2023.

[119] D. Zekri, T. Delot, M. Thilliez, S. Lecomte, and M. Desertot, "A framework for detecting and analyzing behavior changes of elderly people over time using learning techniques," *Sensors (Switzerland)*, vol. 20, no. 24, pp. 1–27, 2020.

[120] J. G. Giménez Manuel, J. C. Augusto, and J. Stewart, “AnAbEL: towards empowering people living with dementia in ambient assisted living,” *Universal Access in the Information Society*, vol. 21, no. 2, pp. 457–476, 2022.

[121] L. Alawneh, M. Al-Ayyoub, Z. A. Al-Sharif, and A. Shatnawi, “Personalized human activity recognition using deep learning and edge-cloud architecture,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 9, pp. 12021–12033, 2022.

[122] N. Gulati and P. D. Kaur, “FriendCare-AAL: a robust social IoT based alert generation system for ambient assisted living,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 4, pp. 1735–1762, 2022.

[123] A. O. Ige and M. H. Mohd Noor, “A survey on unsupervised learning for wearable sensor-based activity recognition,” *Applied Soft Computing*, vol. 127, p. 109363, 2022.

[124] C. Chakraborty and M. R. Khosravi, “Intelligent Healthcare: Infrastructure, Algorithms and Management,” *Intelligent Healthcare: Infrastructure, Algorithms and Management*, no. June, pp. 1–498, 2022.

[125] B. K. Priya, V. S. Koushik Tanniru, and M. Katamaneni, “Ensemble Learning Model for Diabetes Prediction,” in *International Conference on Innovative Data Communication Technologies and Application, ICIDCA 2023 - Proceedings*, pp. 33–36, 2023.

[126] S. Shukralia, N. Delhi, N. Delhi, P. Chakraborty, and N. Delhi, “Fall Detection of Elderly in Ambient Assisted Smart Living Using CNN Based Ensemble Approach,” in *E-business technologies Conference*, pp. 134–139, 2023.

[127] C. P. Liu, J. H. Li, E. P. Chu, C. Y. Hsieh, K. C. Liu, C. T. Chan, and Y. Tsao, “Deep Learning-based Fall Detection Algorithm Using Ensemble Model of Coarse-fine CNN and GRU Networks,” in *IEEE International Symposium on Medical Measurements and Applications*, 2023.

[128] E. A. De La Cal, M. Fáñez, M. Villar, J. R. Villar, and V. M. González, “A low-power HAR method for fall and high-intensity ADLs identification using wrist-worn accelerometer devices,” *Logic Journal of the IGPL*, vol. 31, no. 2, pp. 375–389, 2023.

[129] Y. W. Chen, K. chung Lin, Y. chun Li, and C. J. Lin, “Predicting patient-reported outcome of activities of daily living in stroke rehabilitation: a

machine learning study," *Journal of NeuroEngineering and Rehabilitation*, vol. 20, no. 1, pp. 1–12, 2023.

[130] D. D. Onthoni and P. K. Sahoo, "Artificial-Intelligence-Assisted Activities of Daily Living Recognition for Elderly in Smart Home," *Electronics (Switzerland)*, vol. 11, no. 24, pp. 1–16, 2022.

[131] R. Blanquero, E. Carrizosa, P. Ramírez-Cobo, and M. R. Sillero-Denamiel, "Variable selection for Naïve Bayes classification," *Computers and Operations Research*, vol. 135, p. 105456, 2021.

[132] A. Som, N. Krishnamurthi, M. Buman, and P. Turaga, "Unsupervised Pretrained Models from Healthy ADLs Improve Parkinson's Disease Classification of Gait Patterns," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2020-July, pp. 784–788, 2020.

[133] F. Rustam, A. A. Reshi, I. Ashraf, A. Mehmood, S. Ullah, D. M. Khan, and G. S. Choi, "Sensor-Based Human Activity Recognition Using Deep Stacked Multilayered Perceptron Model," *IEEE Access*, vol. 8, pp. 218898–218910, 2020.

[134] L. N. Kwon, D. H. Yang, M. G. Hwang, S. J. Lim, Y. K. Kim, J. G. Kim, K. H. Cho, H. W. Chun, and K. W. Park, "Automated classification of normal control and early-stage dementia based on activities of daily living (ADL) data acquired from smart home environment," *International Journal of Environmental Research and Public Health*, vol. 18, no. 24, pp. 1–24, 2021.

[135] A. Tousi and M. Lujan, "Comparative Analysis of Machine Learning Models for Performance Prediction of the SPEC Benchmarks," *IEEE Access*, vol. 10, pp. 11994–12011, 2022.

[136] D. R. Beddiar, B. Nini, M. Sabokrou, and A. Hadid, "Vision-based human activity recognition: a survey," *Multimedia Tools and Applications*, vol. 79, no. 41-42, pp. 30509–30555, 2020.

[137] L. Chen, C. D. Nugent, and H. Wang, "A knowledge-driven approach to activity recognition in smart homes," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 6, pp. 961–974, 2012.

[138] A. W. Mohamed, A. A. Hadi, and A. K. Mohamed, "Gaining-sharing knowledge based algorithm for solving optimization problems: a novel nature-inspired algorithm," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 7, pp. 1501–1529, 2020.

[139] J. C. Augusto, J. Liu, P. McCullagh, H. Wang, and J.-B. Yang, "Management of Uncertainty and Spatio-Temporal Aspects for Monitoring and Diagnosis in a Smart Home," *International Journal of Computational Intelligence Systems*, vol. 1, no. 4, p. 361, 2008.

[140] T. Akashi, M. Nakamura, K. Yasuda, and S. Saiki, "Proposal for a Personalized Adaptive Speaker Service to Support the Elderly at Home," in *Proceedings - 22nd IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2021-Fall*, pp. 2–7, IEEE, 2021.

[141] T. G. Stavropoulos, E. Kontopoulos, N. Bassiliades, J. Argyriou, A. Bikakis, D. Vrakas, and I. Vlahavas, "Rule-based approaches for energy savings in an ambient intelligence environment," *Pervasive and Mobile Computing*, vol. 19, pp. 1–23, 2013.

[142] J. Hallberg, C. D. Nugent, R. J. Davies, K. Synnes, M. P. Donnelly, S. Member, D. Finlay, and M. Mulvenna, "HomeRuleML – A Model for the Exchange of Decision Support Rules Within Smart Environments," in *IEEE Automation Science and Engineering*, (Scottsdale), pp. 513–520, 2007.

[143] T. G. Stavropoulos, D. Vrakas, and I. Vlahavas, "A survey of service composition in ambient intelligence environments," *Artificial Intelligence Review*, vol. 40, no. 3, pp. 247–270, 2013.

[144] S. K. Shahzad, D. Ahmed, M. R. Naqvi, M. T. Mushtaq, M. W. Iqbal, and F. Munir, "Ontology Driven Smart Health Service Integration," *Computer Methods and Programs in Biomedicine*, vol. 207, p. 106146, 2021.

[145] T. Gu, X. H. Wang, H. K. Pung, and D. Q. Zhang, "An Ontology-based Context Model in Intelligent Environments," in *arXiv*, 2020.

[146] M. Knorr, *On Combining Ontologies and Rules.* 2022.

[147] A. B. K. A. Babu and R. Sivakumar, "Development of Ontology based Middleware for Context Awareness in Ambient Intelligence," in *Hybrid Intellifent Systems*, (Kuwait), pp. 219–224, IEEE, 2014.

[148] D. Hooda and R. Rani, "Ontology driven human activity recognition in heterogeneous sensor measurements," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 12, pp. 5947–5960, 2020.

[149] X. Chen, S. Jia, and Y. Xiang, "A review: Knowledge reasoning over knowledge graph," *Expert Systems with Applications*, vol. 141, 2020.

[150] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, "An overview of clinical decision support systems: benefits, risks, and strategies for success," *npj Digital Medicine*, vol. 3, no. 1, pp. 1–10, 2020.

[151] R. Sokullu, M. A. Akkaş, and E. Demir, "IoT supported smart home for the elderly," *Internet of Things (Netherlands)*, vol. 11, 2020.

[152] B. N. Schilit, N. Adams, and R. Want, "Context-aware computing applications," in *Proceedings of the 1994 First Workshop on Mobile Computing Systems and Applications*, pp. 85—-90, IEEE Comput. Soc, 1994.

[153] J. Pascoe, "Adding Generic Contextual Capabilities to Wearable Computers," in *International Symposium on Wearable Computers*, vol. 44, pp. 92–99, IEEE, 1998.

[154] E. Stip and V. Rialle, "Environmental Cognitive Remediation of Schizophrenia: Ethical Implications of "Smart Home" Technology," *The Canadian Journal of Psychiatry*, vol. 50, no. 5, pp. 281–291, 2005.

[155] M. Seliem, K. Elgazzar, and K. Khalil, "Towards Privacy Preserving IoT Environments : A Survey," *Wireless Communications and Mobile Computing*, vol. 2018, no. October 2016, 2018.

[156] Z. Guan, Y. Zhang, L. Wu, J. Wu, J. Li, Y. Ma, and J. Hu, "APPA: An anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT," *Journal of Network and Computer Applications*, vol. 125, no. August 2018, pp. 82–92, 2019.

[157] M. Memon, S. R. Wagner, C. F. Pedersen, F. H. Aysha Beevi, and F. O. Hansen, "Ambient Assisted Living healthcare frameworks, platforms, standards, and quality attributes," *Sensors (Switzerland)*, vol. 14, no. 3, pp. 4312–4341, 2014.

[158] G. Yee, L. Korba, and R. Song, "Ensuring privacy for E-health services," in *Proceedings - First International Conference on Availability, Reliability and Security, ARES 2006*, vol. 2006, pp. 321–328, 2006.

[159] A. Pfitzmann and K. Borcea-Pfitzmann, "Lifelong Privacy: Privacy and Identity Management for Life," in *IFIP Advances in Information and Communication Technology*, pp. 1–17, 2010.

[160] J. Hightower and G. Borriello, "Location Systems for Ubiquitous Computing," *Computer*, vol. 34, no. 8, pp. 57–66, 2001.

[161] H. Najeh, C. Lohr, and B. Leduc, "Convolutional Neural Network Bootstrapped by Dynamic Segmentation and Stigmergy-Based Encoding for Real-Time Human Activity Recognition in Smart Homes," *Sensors*, vol. 23, no. 4, 2023.

[162] M. Webber and R. F. Rojas, "Human Activity Recognition with Accelerometer and Gyroscope: A Data Fusion Approach," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 16979–16989, 2021.

[163] G. Fischer, J. Bordoy, D. J. Schott, W. Xiong, A. Gabbrielli, F. Hoflinger, K. Fischer, C. Schindelhauer, and S. J. Rupitsch, "Multimodal Indoor Localization: Fusion Possibilities of Ultrasonic and Bluetooth Low-Energy Data," *IEEE Sensors Journal*, vol. 22, no. 6, pp. 5857–5868, 2022.

[164] H. Pirsiavash and D. Ramanan, "Detecting activities of daily living in first-person camera views," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, (Providence, RI), pp. 2847–2854, Ieee, jun 2012.

[165] B. Aasand and C. H. Min, "Automatic Recognition of Activities of Daily Living for Elderly to Support Independent Living," *2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2022*, pp. 286–290, 2022.

[166] G. Xu, Q. Wan, W. Deng, T. Guo, and J. Cheng, "Smart-Sleeve: A Wearable Textile Pressure Sensor Array for Human Activity Recognition," *Sensors*, vol. 22, no. 5, pp. 1–17, 2022.

[167] Google, "Google Glass," 2014, support.google.com/glass/answer/3064128, (accessed May. 13, 2014).

[168] Google, “Google Glass Enterprise,” 2022, https://www.google.com/glass/, (accessed Oct. 2, 2022).

[169] Vuzix, “M100, Smart Glasses,” 2014, http://www.vuzix.com/, (accessed May. 13, 2014).

[170] H. Khan, A. Clark, G. Woodward, and R. W. Lindeman, “Improved position accuracy of foot-mounted inertial sensor by discrete corrections from vision-based fiducial marker tracking,” *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–16, 2020.

[171] G. M. Mendoza-Silva, J. Torres-Sospedra, and J. Huerta, “A meta-review of indoor positioning systems,” *Sensors (Switzerland)*, vol. 19, no. 20, 2019.

[172] Y. Rahal, H. Pigot, and P. Mabilleau, “Location estimation in a smart home: System implementation and evaluation using experimental data,” *International Journal of Telemedicine and Applications*, vol. 2008, no. 4, p. 9, 2008.

[173] G. Okeyo, L. Chen, and H. Wang, “An Agent-mediated Ontology-based Approach for Composite Activity Recognition in Smart Homes,” *Journal of Universal Computer Science*, vol. 19, no. 17, pp. 2577–2597, 2013.

[174] L. Kanaris, A. Kokkinis, A. Liotta, and S. Stavrou, “Fusing bluetooth beacon data with Wi-Fi radiomaps for improved indoor localization,” *Sensors (Switzerland)*, vol. 17, no. 4, pp. 1–15, 2017.

[175] O. Bin Tariq, M. T. Lazarescu, J. Iqbal, and L. Lavagno, “Performance of Machine Learning Classifiers for Indoor Person Localization with Capacitive Sensors,” *IEEE Access*, vol. 5, pp. 12913–12926, 2017.

[176] F. Antoniazzi, G. Paolini, L. Roffia, D. Masotti, A. Costanzo, and T. S. Cinotti, “A web of things approach for indoor position monitoring of elderly and impaired people,” in *Conference of Open Innovation Association, FRUCT*, pp. 51–56, 2017.

[177] A. R. Jiménez, F. Seco, P. Peltola, and M. Espinilla, “Location of Persons Using Binary Sensors and BLE Beacons for Ambient Assitive Living,” in *IPIN 2018 - 9th International Conference on Indoor Positioning and Indoor Navigation*, no. September, pp. 24–27, IEEE, 2018.

[178] S. A. Maghdid, H. S. Maghdid, and S. Rahman, “Tracking Indoor Elderly-People Using Onboard Smartphones Wi-Fi Device and Inertial Sensors,” in

1st International Conference on Advanced Research in Engineering Sciences, ARES 2018, IEEE, 2018.

[179] V. Bianchi, P. Ciampolini, and I. De Munari, "RSSI-Based Indoor Localization and Identification for ZigBee Wireless Sensor Networks in Smart Homes," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 2, pp. 566–575, 2019.

[180] M. Kolakowski, "Improving accuracy and reliability of Bluetooth Low-Energy-based localization systems using proximity sensors," *Applied Sciences (Switzerland)*, vol. 9, no. 19, 2019.

[181] E. S. Sansano, O. Belmonte-Fernandez, R. Montoliu, A. Gasco-Compte, A. C. Miedes, and P. B. Iturralde, "Improving positioning accuracy in ambient assisted living environments. a multi-sensor approach," in *Proceedings - 2019 15th International Conference on Intelligent Environments, IE 2019*, pp. 22–29, 2019.

[182] A. V. Vesa, S. Vlad, R. Rus, M. Antal, C. Pop, I. Anghel, T. Cioara, and I. Salomie, "Human Activity Recognition using Smartphone Sensors and Beacon-based Indoor Localization for Ambient Assisted Living Systems," in *Proceedings - 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing, ICCP 2020*, pp. 205–212, 2020.

[183] J. Kolakowski, V. Djaja-Josko, M. Kolakowski, and J. Cichocki, "Localization system supporting people with cognitive impairment and their caregivers," *International Journal of Electronics and Telecommunications*, vol. 66, no. 1, pp. 125–131, 2020.

[184] A. Bilbao-Jayo, A. Almeida, I. Sergi, T. Montanaro, L. Fasano, M. Emaldi, and L. Patrono, "Behavior modeling for a beacon-based indoor location system," *Sensors*, vol. 21, no. 14, pp. 1–15, 2021.

[185] J. D. Ceron, D. M. López, F. Kluge, and B. M. Eskofier, "Framework for Simultaneous Indoor Localization, Mapping, and Human Activity Recognition in Ambient Assisted Living Scenarios," *Sensors*, vol. 22, no. 9, 2022.

[186] S. Parmar, X. Wang, C. Yang, and S. Mao, "Voice Fingerprinting for Indoor Localization with a Single Microphone Array and Deep Learning," in *WiseML 2022 - Proceedings of the 2022 ACM Workshop on Wireless Security and Machine Learning*, pp. 21–26, 2022.

[187] E. Ackerman, "Google Gets in Your Face," *IEEE Spectrum*, vol. 50, no. 1, pp. 26–29, 2013.

[188] F. Leotta and M. Mecella, "PLaTHEA: a marker-less people localization and tracking system for home automation," *Software - Practice and Experience*, vol. 39, no. 7, pp. 661–699, 2014.

[189] P. I. Wilson and J. Fernandez, "Facial Feature Detection Using Haar Classifiers," *Journal of Computing Sciences in Colleges*, vol. 21, no. 4, pp. 127–133, 2006.

[190] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[191] A. Zeb, S. Ullah, and I. Rabbi, "Indoor Vision-Based Auditory Assistance for Blind People in Semi Controlled Environments," in *Image Processing Theory, Tools and Applications*, pp. 1–6, 2014.

[192] J. Rivera-rubio, I. Alexiou, A. Bharath, R. Secoli, L. Dickens, and E. C. Lupu, "Associating locations from wearable cameras," in *British Machine Vision Conference*, pp. 1–13, 2014.

[193] D. Zhang, D. J. Lee, and B. Taylor, "Seeing Eye Phone: A smart phone-based indoor localization and guidance system for the visually impaired," *Machine Vision and Applications*, vol. 25, no. 3, pp. 811–822, 2014.

[194] C. Orrite, J. Soler, M. Rodríguez, E. Herrero, and R. Casas, "Image-based Location Recognition and Scenario Modelling," in *International Conference on Computer Vision Theory and Applications*, pp. 216–221, 2015.

[195] W. Chojnacki, M. J. Brooks, A. V. D. Hengel, and D. Gawley, "Revisiting Hartley ' s Normalized Eight-Point Algorithm," *IEEE transactions on pattern analysis and machine intelligence*, vol. 25, no. 9, pp. 1172–1177, 2003.

[196] M. J. Edwards, M. P. Hayes, and R. D. Green, "High-accuracy Fiducial Markers for Ground Truth," in *Image and Vision Computing New Zealand*, IEEE, 2016.

[197] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[198] A. Rituerto, G. Fusco, and J. M. Coughlan, “Towards a Sign-Based Indoor Navigation System for People with Visual Impairments,” in *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS ’16*, pp. 287–288, ACM, 2016.

[199] G. Kapidis, R. W. Poppe, E. A. Van Dam, R. C. Veltkamp, and L. P. Noldus, “Where Am I? Comparing CNN and LSTM for Location Classification in Egocentric Videos,” in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2018*, pp. 878–883, IEEE, 2018.

[200] J. Redmon, “Darknet: Open Source Neural Networks in C,” 2016, (accessed Oct. 2, 2023).

[201] J. D. Domingo, J. Gómez-García-Bermejo, E. Zalama, C. Cerrada, and E. Valero, “Integration of computer vision and wireless networks to provide indoor positioning,” *Sensors (Switzerland)*, vol. 19, no. 24, pp. 1–17, 2019.

[202] E. Martin-Gorostiza, M. A. Garcia-Garrido, D. Pizarro, P. Torres, M. O. Miguel, and D. Salido-Monzu, “Infrared and camera fusion sensor for indoor positioning,” in *2019 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2019*, 2019.

[203] F. M. Li, D. L. Chen, M. Fan, and K. Truong, “FMT: A Wearable Camera-Based Object Tracking Memory Aid for Older Adults,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–25, 2019.

[204] W. Hu, K. Wang, H. Chen, R. Cheng, and K. Yang, “An indoor positioning framework based on panoramic visual odometry for visually impaired people,” *Measurement Science and Technology*, vol. 31, no. 1, 2019.

[205] J. Kunhoth, A. Karkar, S. Al-Maadeed, and A. Al-Attiyah, “Comparative analysis of computer-vision and BLE technology based indoor navigation systems for people with visual impairments,” *International Journal of Health Geographics*, vol. 18, no. 1, 2019.

[206] J. M. Quero, F. Cruciani, L. Seidenari, M. Espinilla, and C. Nugent, “Straightforward Recognition of Daily Objects in Smart Environments from

Wearable Vision Sensor," in *2019 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2019*, pp. 46–51, IEEE, 2019.

[207] D. C. Luvizon, D. Picard, and H. Tabia, "2D/3D pose estimation and action recognition using multitask deep learning," in *arXiv*, 2018.

[208] I. Uygur, R. Miyagusuku, S. Pathak, A. Moro, A. Yamashita, and H. Asama, "Robust and efficient indoor localization using sparse semantic information from a spherical camera," *Sensors (Switzerland)*, vol. 20, no. 15, pp. 1–21, 2020.

[209] M. Košťák and A. Slabý, "Designing a simple fiducial marker for localization in spatial scenes using neural networks," *Sensors*, vol. 21, no. 16, 2021.

[210] S. Li, B. Yu, Y. Jin, L. Huang, H. Zhang, and X. Liang, "Image-Based Indoor Localization Using Smartphone Camera," *Wireless Communications and Mobile Computing*, vol. 2021, 2021.

[211] T. Zhou, J. Ku, B. Lian, and Y. Zhang, "Indoor positioning algorithm based on improved convolutional neural network," *Neural Computing and Applications*, vol. 34, no. 9, pp. 6787–6798, 2022.

[212] M. Tabuchi and T. Hirotomi, "Using Fiducial Marker for Analyzing Wearable Eye-Tracker Gaze Data Measured While Cooking," in *International Conference on Human-Computer Interaction*, pp. 192 – 204, Springer, 2022.

[213] M. Afzalan and F. Jazizadeh, "Indoor positioning based on visible light communication: A performance-based survey of real-world prototypes," *ACM Computing Surveys*, vol. 52, no. 2, 2019.

[214] P. Giannakeris, K. Avgerinakis, S. Vrochidis, and I. Kompatsiaris, "Activity recognition from wearable cameras," in *Proceedings - International Workshop on Content-Based Multimedia Indexing*, vol. 2018-Septe, IEEE, 2018.

[215] S. Noor and V. Uddin, "First Person Vision for Activity Prediction Using Probabilistic Modeling," *Mehran University Research Journal of Engineering and Technology*, vol. 37, no. 4, pp. 545–558, 2018.

[216] M. Tenorth, J. Bandouch, and M. Beetz, "The TUM Kitchen Data Set of Everyday Manipulation Activities for Motion Tracking and Action Recognition," in *IEEE International Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS)*, 2009.

[217] Y. Li, Z. Ye, and J. M. Rehg, “Delving into egocentric actions,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June, no. 1, pp. 287–295, 2015.

[218] Z. Zuo, L. Yang, Y. Peng, F. Chao, and Y. Qu, “Gaze-informed egocentric action recognition for memory aid systems,” *IEEE Access*, vol. 6, pp. 12894–12904, 2018.

[219] Tobii Ltd., “Tobii Pro Glasses 2.” (accessed May. 12, 2014).

[220] H. Yu, G. Pan, M. Pan, C. Li, W. Jia, L. Zhang, and M. Sun, “A hierarchical deep fusion framework for egocentric activity recognition using a wearable hybrid sensor system,” *Sensors (Switzerland)*, vol. 19, no. 3, 2019.

[221] M. Sun, L. Burke, Z.-H. Mao, Y. Chen, H.-C. Chen, Y. Bai, Y. Li, C. Li, and W. Jia, “eButton: A Wearable Computer for Health Monitoring and Personal Assistance,” in *Design Automatinon Conference*, vol. 23, pp. 1–7, 2014.

[222] S. Song, V. Chandrasekhar, B. Mandal, L. Li, J. H. Lim, G. S. Babu, P. P. San, and N. M. Cheung, “Multimodal Multi-Stream Deep Learning for Egocentric Activity Recognition,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 378–385, IEEE, 2016.

[223] A. Diete and H. Stuckenschmidt, “Fusing object information and inertial data for activity recognition,” *Sensors (Switzerland)*, vol. 19, no. 19, 2019.

[224] F. De la Torre, J. Hodgins, J. Montano, S. Valcarcel, and J. Macey, “Guide to the Carnegie Mellon University multimodal activity (CMU-MMAC) database,” 2009.

[225] B. X. B. Yu, Y. Liu, and K. C. C. Chan, “Vision-Based Daily Routine Recognition for Healthcare with Transfer Learning,” *International Journal of Biomedical and Biological Engineering*, vol. 14, no. 7, pp. 178–186, 2020.

[226] J. Massardi, M. Gravel, and E. Beaudry, “PARC: A Plan and Activity Recognition Component for Assistive Robots,” in *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 3025–3031, 2020.

[227] M. Su, D. W. Hayati, S. Tseng, J. Chen, and H. Wei, “Smart care using a dnn-based approach for activities of daily living (ADL) recognition,” *Applied Sciences (Switzerland)*, vol. 11, no. 1, pp. 1–12, 2021.

[228] J. Joyce, "Bayes' Theorem," in *Stanford Encyclopedia of Philosophy Archive* (E. N. Zalta, ed.), ch. Bayes' The, Metaphysics Research Lab, Stanford University, 2019.

[229] D. Berrar, "Bayes' Theorem and Naive Bayes Classifier," in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, ch. Bayes' The, pp. 403 – 412, 2019.

[230] G. K. Ramadhan and D. N. Utama, "Fuzzy Tsukamoto based Decision Support Model for Purchase Decision in Pharmacy Company," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 4, pp. 3868–3874, 2019.

[231] M. D. Sinaga, N. S. Br Sembiring, E. Darmayanti, R. Maulida, E. V. Haryanto, and E. Ginting, "A Comparison of the Tsukamoto Fuzzy Method and the Dempster Shafer Method for Diagnosing Leptospirosis," *2021 9th International Conference on Cyber and IT Service Management, CITSM 2021*, pp. 13–17, 2021.

[232] D. Ayres, J. Schmutte, and J. Stanfield, "Expect the unexpected: Risk assessment using Monte Carlo simulations," *Journal of Accountancy*, 2017.

[233] R. A. Sharif and S. Pokharel, "Risk analysis with the dempster–shafer theory for smart city planning: The case of qatar," *Electronics (Switzerland)*, vol. 10, no. 24, 2021.

[234] T. Bhowmik, R. Mojumder, D. Ghosh, and I. Banerjee, "IoT Based Smart Home-Health Monitoring System Using Dempster-Shafer Evidence Theory for Pandemic Situation," in *ACM International Conference Proceeding Series*, no. 1, pp. 260–265, 2022.

[235] G. Shafer, "The Dempster-Shafer Theory." 2011.

[236] K. Sentz and S. Ferson, "Combination of Evidence in Dempster- Shafer Theory," *Contract*, no. April, p. 96, 2002.

[237] S. Razi, M. R. Karami Mollaei, and J. Ghasemi, "A novel method for classification of BCI multi-class motor imagery task based on Dempster–Shafer theory," *Information Sciences*, vol. 484, pp. 14–26, 2019.

[238] J. M. Alcalá, J. Ureña, Á. Hernández, and D. Gualda, "Assessing human activity in elderly people using non-intrusive load monitoring," *Sensors (Switzerland)*, vol. 17, no. 2, 2017.

[239] L. Zhao and G. Li, "Abnomal Activity Identification Method Based on Ontology and Dempster-Shafer Theory," in *Proceedings - 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics, IHMSC 2018*, vol. 1, pp. 289–292, IEEE, 2018.

[240] F. Al Machot, H. C. Mayr, and S. Ranasinghe, "A hybrid reasoning approach for activity recognition based on answer set programming and dempster–shafer theory," *Studies in Systems, Decision and Control*, vol. 109, pp. 303–318, 2018.

[241] H. Sfar and A. Bouzeghoub, "Activity Recognition for Anomalous Situations Detection," *IRBM*, vol. 39, no. 6, pp. 400–406, 2018.

[242] V. Venkatesh, P. Raj, K. Kannan, and P. Balakrishnan, "Precision centric framework for activity recognition using Dempster Shaffer theory and information fusion algorithm in smart environment," *Journal of Intelligent and Fuzzy Systems*, vol. 36, no. 3, pp. 2117–2124, 2019.

[243] N. Mishra, S. Chaturvedi, A. Vij, and S. Tripathi, "Research Problems in Recommender systems," *Journal of Physics: Conference Series*, vol. 1717, p. 012002, 2021.

[244] H. Pirsiavash and D. Ramanan, "Detecting Activities of Daily Living in First-person Camera Views.pdf," in *Computer Vision and Pattern Recognition*, IEEE, 2012.

[245] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "HMDB: A large video database for human motion recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2556–2563, IEEE, 2011.

[246] K. Soomro, A. R. Zamir, and M. Shah, "UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild," in *arXiv*, no. November, 2012.

[247] W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev, M. Suleyman, and A. Zisserman, "The Kinetics human action video dataset," *arXiv*, 2017.

[248] G. A. Sigurdsson, G. Varol, X. Wang, A. Farhadi, I. Laptev, and A. Gupta, "Hollywood in homes: Crowdsourcing data collection for activity understanding," in *Lecture Notes in Computer Science (including subseries Lec-*

ture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9905 LNCS, pp. 510–526, 2016.

[249] A. Shahroudy, J. Liu, T. T. Ng, and G. Wang, "NTU RGB+D: A large scale dataset for 3D human activity analysis," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 1010–1019, IEEE, 2016.

[250] D. Weinland, R. Ronfard, E. Boyer, D. Weinland, R. Ronfard, E. Boyer, D. Weinland, R. Ronfard, and E. Boyer, "Free viewpoint action recognition using motion history volumes To cite this version : HAL Id : inria-00544629 Free Viewpoint Action Recognition using Motion History Volumes," *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 249–257, 2011.

[251] M. D. Rodriguez, J. Ahmed, and M. Shah, "Action MACH: A spatio-temporal maximum average correlation height filter for action recognition," in *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008.

[252] M. Marszałek, I. Laptev, and C. Schmid, "Actions in context," in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, vol. 2009 IEEE, pp. 2929–2936, IEEE, 2009.

[253] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1290–1297, IEEE, 2012.

[254] K. Reddy and M. Shah, "Recognizing 50 human action categories of web videos," in *Machine Vision and Applications*, vol. 24, pp. 971–981, 2013.

[255] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and F. F. Li, "Large-scale video classification with convolutional neural networks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1725–1732, 2014.

[256] P. Weinzaepfel, X. Martin, and C. Schmid, "Human Action Localization with Sparse Spatial Supervision," in *arXiv*, pp. 1–14, 2016.

[257] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, 2022.

[258] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, no. November, p. 100258, 2022.

[259] S. Giroux, M. Castebrunet, O. Boissier, and V. Rialle, "A Multiagent Approach to Personalization and Assistance to Multiple Persons in a Smart Home," in *Artificial Intelligence Applied to Assistive Technologies and Smart Environments*, no. Artifical Intelligence Applied to Assistive Technologies and Smart Environments, pp. 11–15, 2014.

[260] Q. Li, R. Gravina, Y. Li, S. H. Alsamhi, F. Sun, and G. Fortino, "Multiuser activity recognition: Challenges and opportunities," *Information Fusion*, vol. 63, no. March, pp. 121–135, 2020.

[261] A. Lotfi, C. Langensiepen, S. M. Mahmoud, and M. J. 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.

[262] B. Brumitt, B. Meyers, J. Krumm, A. Kern, and S. Shafer, *Handheld and Ubiquitous Computing*. Springer Berlin Heidelberg, 2000.

[263] F. Zafari, A. Gkelias, and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.

[264] C. Shewell, C. Nugent, and H. Wang, "Indoor localisation through object detection within multiple environments utilising a single wearable camera," *Health and Technology*, vol. 7, no. 1, pp. 51–60, 2017.

[265] H. Pirsiavash and D. Ramanan, "Detecting activities of daily living in first-person camera views," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2847–2854, IEEE, 2012.

[266] D. Dua and C. Graff, "UCI Machine Learning Repository," 2019, (accessed Oct. 2, 2023).

[267] A. Borreo, L. Onofri, and P. Soda, "A multi-environment dataset for activity of daily living recognition in video streams," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (Milan, Italy), IEEE, 2015.

[268] J. Wang, X. Nie, Y. Xia, Y. Wu, and S. C. Zhu, "Cross-view action modeling, learning, and recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2649–2656, 2014.

[269] H. Rahmani, A. Mahmood, D. Huynh, and A. Mian, "Histogram of Oriented Principal Components for Cross-View Action Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 12, pp. 2430–2443, 2016.

[270] P. Edemekong, D. Bomgaars, S. Sukumaran, and C. Schoo, "Activities of Daily Living," 2022, (accessed Nov. 29, 2022).

[271] C. Nugent, M. Mulvenna, X. Hong, and S. Devlin, "Experiences in the development of a Smart Lab," *International Journal of Biomedical Engineering and Technology*, vol. 2, no. 4, p. 319, 2009.

[272] M. Espinilla, L. Martinez, J. Medina, and C. Nugent, "The experience of developing the UJAmI Smart Lab," *IEEE Access*, vol. 6, pp. 34631–34642, 2018.

[273] TyneTec, "http://www.tynetec.co.uk/," 2017, (accessed Nov. 22, 2017).

[274] R. LiKamWa, Z. Wang, A. Carroll, F. X. Lin, and L. Zhong, "Draining our glass," in *Proceedings of 5th Asia-Pacific Workshop on Systems*, pp. 1–7, ACM, 2014.

[275] Oracle, "http://www.sunspotdev.org/," 2017, http://www.sunspotdev.org/, (accessed Nov. 22, 2017).

[276] Philips, "https://www.lighting.philips.co.uk," 2020, https://www.lighting.philips.co.uk/, (accessed Dec. 7, 2020).

[277] J. King, E. Jansen, S. Helal, W. Mann, H. El-Zabadani, and Y. Kaddoura, "The Gator Tech Smart House : a programmable pervasive space," *Computer*, vol. 38, no. 3, pp. 64–74, 2005.

[278] D. J. Cook, A. Crandall, B. Thomas, and N. Krishnan, "CASAS: A Smart Home in a Box," *Computer*, vol. 46, no. 7, pp. 62–69, 2013.

[279] D. Cook, M. Youngblood, E. Heierman, K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja, "MavHome: an agent-based smart home," in *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003. (PerCom 2003).*, pp. 521–524, IEEE Comput. Soc, 2003.

[280] J. A. Kientz and B. Jones, "The Georgia Tech Aware Home," in *Human Factors in Computing Systems*, pp. 3675–3680, 2008.

[281] K. Synnes, M. Lilja, A. Nyman, M. Espinilla, I. Cleland, A. G. S. Comas, Z. Comas-Gonzalez, J. Hallberg, N. Karvonen, W. O. de Morais, F. Cruciani, and C. Nugent, "H2Al—The Human Health and Activity Laboratory," in *MDPI Proceedings*, vol. 2, p. 1241, 2018.

[282] S. Fallmann, I. Psychoula, L. Chen, F. Chen, J. Doyle, and D. Triboan, "Reality and perception: Activity monitoring and data collection within a real-world smart home," in *2017 IEEE SmartWorld Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation, SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI 2017* -, pp. 1–6, IEEE, 2018.

[283] S. Mennicken and E. M. Huang, "Hacking the natural habitat: An in-the-wild study of smart homes, their development, and the people who live in them," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7319 LNCS, no. June, pp. 143–160, 2012.

[284] J. Synnott, C. Nugent, S. Zhang, A. Calzada, I. Cleland, M. Espinilla, J. M. Quero, and J. Lundstrom, "Environment Simulation for the Promotion of the Open Data Initiative," in *2016 IEEE International Conference on Smart Computing, SMARTCOMP 2016*, 2016.

[285] C. Shewell, C. Nugent, M. Donnelly, and H. Wang, "Indoor Localisation Through Object Detection on Real-Time Video Implementing a Single Wearable Camera," in *Mediterranean Conference on Medical and Biological Engineering and Computing* (E. Kyriacou, S. Christofides, and C. Pattichis,

eds.), (Paphos, Cyprus), pp. 1231–1236, Springer International Publishing, 2016.

[286] J. Iglesias, A. M. Bernardos, and J. R. Casar, "A Bayesian Strategy to Enhance the Performance of Indoor Localization Systems," in *Advances in Intelligent and Soft Computing* (J. M. Molina, J. R. C. Corredera, A. M. B. Barbolla, M. F. C. Perez, and J. Ortega-Garcia, eds.), vol. 94, pp. 79–92, 2011.

[287] SERG, "Lab Resources," 2012, http://scm.ulster.ac.uk/, (accessed Jan. 1, 2017).

[288] M. Fiala, "Designing highly reliable fiducial markers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1317–1324, 2010.

[289] C. Owen, F. X. F. Xiao, and P. Middlin, "What is the best fiducial?," *The First IEEE International Workshop Agumented Reality Toolkit,*, vol. 15, no. 11, p. 3317, 2002.

[290] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 105–119, 2010.

[291] A. Sagitov, K. Shabalina, R. Lavrenov, and E. Magid, "Comparing fiducial marker systems in the presence of occlusion," in *2017 International Conference on Mechanical, System and Control Engineering, ICMSC 2017*, pp. 377–382, IEEE, 2017.

[292] OpenCV, "OpenCV," 2020, https://opencv.org/, (accessed Oct. 2, 2023).

[293] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *International Conference on Computer Vision*, (Barcelona), pp. 2564–2571, IEEE, 2011.

[294] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[295] M. Fanqing and Y. Funcheng, "A Tracking Algorithm Based on ORB," in *International Conference on Mechatronic Sciences, Electric Engineering and Computer*, no. 1, (Shengyang), pp. 1187–1190, IEEE, 2013.

[296] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, nov 2004.

[297] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3951 LNCS, pp. 404–417, 2006.

[298] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3951 LNCS, pp. 430–443, 2006.

[299] P. Rosin, "Measuring Corner Properties," *Computer Vision and Image Understanding*, vol. 73, no. 2, pp. 291–307, 1999.

[300] A. Gil, O. M. Mozos, M. Ballesta, and O. Reinoso, "A comparative evaluation of interest point detectors and local descriptors for visual SLAM," *Machine Vision and Applications*, vol. 21, no. 6, pp. 905–920, 2010.

[301] OpenCV, "Feature Matching," 2014, (accessed Oct. 2, 2023).

[302] D. Verma, N. Kakkar, and N. Mehan, "Comparison of Brute-Force and K-D Tree Algorithm," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 3, no. 1, 2014.

[303] J. Cheng, C. Leng, J. Wu, H. Cui, and H. Lu, "Fast and Accurate Image Matching with Cascade Hashing for 3D Reconstruction," in *Computer Vision and Pattern Recognition*, (Columbus, OH), pp. 1–8, IEEE Comput. Soc, 2014.

[304] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6314 LNCS, no. PART 4, pp. 778–792, 2010.

[305] The College of Optometrists, "Britain's Eye Health in Focus." 2013.

[306] C. Shewell, C. Nugent, M. Donnelly, and H. Wang, "Wearable Computing to Support Activities of Daily Living," in *International Workconference on Ambient Assisted Living* (L. Pecchia, L. L. Chen, C. Nugent, and J. Bravo, eds.), (Belfast), pp. 195–202, Springer International Publishing, 2014.

[307] S. Garrido-Jurado, R. Muaoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnicer, "Generation of fiducial marker dictionaries using Mixed Integer Linear Programming," *Pattern Recognition*, 2015.

[308] R. G. Brown, "Exponential Smoothing for Predicting Demand," *Cambridge*, 1956.

[309] C. C. Holt, "Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages," *International Journal of Forecasting*, vol. 20, no. 1, pp. 5–10, 2004.

[310] E. Masry, "Alpha-Stable Signals and Adaptive Filtering," *Signal Processing*, vol. 48, no. 11, pp. 3011–3016, 2000.

[311] E. Brookner, "Tracking and Kalman Filtering Made Easy," 1998.

[312] L. A. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[313] J. M. Mendel, "Uncertainty, Fuzzy Logic, and Signal Processing," *Signal Processing*, vol. 80, no. 6, pp. 913–933, 2000.

[314] T. H. C. Nguyen, J. C. Nebel, and F. Florez-Revuelta, "Recognition of activities of daily living with egocentric vision: A review," *Sensors (Switzerland)*, vol. 16, no. 1, 2016.

[315] B. Xiong, G. Kim, and L. Sigal, "Storyline Representation of Egocentric Videos With an Applications to Story-Based Search," in *IEEE International Conference on Computer Vision*, pp. 4525–4533, 2015.

[316] PTC, "Vuforia: Market-Leading Enterprise AR," 2023, (accessed Oct. 4, 2023).

[317] J. Wang and E. Olson, "AprilTag 2: Efficient and robust fiducial detection," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2016-Novem, pp. 4193–4198, 2016.

[318] M. Fiala, "ARTag, a fiducial marker system using digital techniques," *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, vol. II, pp. 590–596, 2005.

[319] D. Fox, J. Hightower, L. Liao, and D. Schulz, "Bayesian Filtering for Location Estimation," *Pervasive Computing, IEEE*, vol. 2, no. 3, pp. 24–33, 2003.

[320] Z. Tong, P. Xu, and T. Denœux, "An evidential classifier based on Dempster-Shafer theory and deep learning," *Neurocomputing*, vol. 450, pp. 275–293, 2021.

[321] L. Li, C. Li, X. Lu, H. Wang, and D. Zhou, "Multi-Focus Image Fusion with Convolutional Neural Network Based on Dempster-Shafer Theory," *Optik*, vol. 272, no. September 2022, p. 170223, 2022.

[322] G. Shafer, "A mathematical theory of evidence.," *Princeton University Press*, 1976.

[323] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, "Machine learning algorithm validation with a limited sample size," *PLoS ONE*, vol. 14, no. 11, pp. 1–20, 2019.

[324] D. J. Hand and K. Yu, "Idiot's Bayes - Not so stupid after all?," *International Statistical Review*, vol. 69, no. 3, pp. 385–398, 2001.

[325] S. Ranganathan, M. Gribskov, K. Nakai, and C. Schonbach, "Encyclopedia of Bioinformatics and Computational Biology," in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, pp. 403 – 412, Elsevier, 2018.

[326] L. Breiman, "A Data Mining Based System for Transaction Fraud Detection," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[327] H. Ramchoun, M. Amine, J. Idrissi, Y. Ghanou, and M. Ettaouil, "Multilayer Perceptron: Architecture Optimization and Training," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 1, p. 26, 2016.

[328] M. Z. Khan, O. H. Alhazmi, M. A. Javed, H. Ghandorh, and K. S. Aloufi, "Reliable internet of things: Challenges and future trends," *Electronics (Switzerland)*, vol. 10, no. 19, pp. 1–22, 2021.

[329] S. McDonald, D. Towey, and V. Brusic, "Social Impact of Smart Environments: Software Engineering Perspectives and Challenges," in *Proceedings - 2022 IEEE 46th Annual Computers, Software, and Applications Conference, COMPSAC 2022*, pp. 1592–1597, IEEE, 2022.

[330] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," *Neural and Evolutionary Computing*, pp. 1–11, 2015.

[331] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *Computer Vision and Pattern Recognition*, 2020.