VIDEO-BASED ESTIMATION OF ACTIVITY LEVEL FOR ASSISTED LIVING

submitted by

Sandipan Pal

for the degree of

Doctor of Philosophy

of the

Department of Electronic and Electrical Engineering
The University of Sheffield

December, 2017

COPYRIGHT

Signature of Author ....

Sandipan Pal

Abstract

The continual increase in the population of older adults in the next 50 *years envisages an increase of dependants on the family and the Government. Assisted Living technologies are information and communication technologies to assist, improve and monitor the daily living of the old and vulnerable population by promoting greater independence and providing a safe and secure environment at a reduced cost. Most of the assisted living technologies are passive sensor-based solutions where a number of embedded or body-worn sensors are employed or connected over a network to recognize activities. Often the sensors are obtrusive and are extremely sensitive to the performance of the sensors. Visual data is contextually richer than sensor triggered firings. Visual data along with being contextual is also extremely sensitive.*

In this work, a camera-based solution for assisted living is proposed. Since visual data is intrusive, a qualitative study among older adults within the community was carried out to get a context of the privacy concerns of having a camera within an assisted living environment. Building on the outcomes of the focus group discussions, a novel monitoring framework is proposed. Following the framework, Activity Level, as an effective metric to measure the amount of activity undertaken by an individual is proposed. Activity Level is estimated by extracting and classifying pixel-based and phase-based motion features. Experiments reveal that phase-based features perform better than pixel-based features. Experiments are carried out using the novel Sheffield Activities of Daily Living Dataset, which has been developed and made available for further computer vision research for assisted living.

To Mummy and Daddy

Acknowledgement

I would like to express my sincere gratitude and thanks to my supervisor Dr. Charith Abhayaratne for guiding me through the long, solitary journey named PhD. Along with being a supervisor, Dr. Abhayaratne has always extended the extra support needed for a PhD student and has always emphasized on the integrity of a PhD degree and being a part of the wider research community. I would also like to thank my second supervisor, Prof. Mark Hawley for supervising me and providing me the opportunity to work on several digital healthcare projects during my time in Sheffield. Working on those projects in parallel with my PhD opened up my thinking avenues, gave me job experience and also provided me with the much needed financial support.

I would like to express my thanks to The University of Sheffield cross-cutting research (Digital World) studentship within the PIPIN network [headed by Prof. Gail Mountain] for funding this research in the first 3 years. I would also like to thank my examiners, Dr. Mohammed Bennaissa and Dr. Dimitrios Makris for giving valuable feedback to the thesis. I would like to express my gratitude to all the volunteers of my focus group study and also to Centre for Assistive Living and Connected Healthcare (CATCH) for providing with the Home Lab. My interactions with my colleagues and students at The University of Sheffield International College [USIC] have enriched me as a person and also made me appreciate all the subtle things that my teachers have done for me over the years.

Sheffield has been my home away from home for the longest period of time till now and this city has given me a set of lovely friends. I would like to specially thank Dr. Arunangshu Chatterjee and Dr. Deepayan Bhowmik for their guidance, encouragement and advice. I would also like to thank my self-styled gang, 'Panchayat' (Ayan, Pallavi, Uday and Sanket) for all those evenings after a tough supervisory meeting reflecting on life and philosophy. I would also like to thank Mr. and Mrs. S. Basu and their lovely children Brishti and Tithi for accommodating me in their house in the last phase of the writing-up period. There are many more whom I couldn't name and to all of them I will always be grateful for their love and kindness. Now that I do not live in Sheffield any more, I do miss the city and its people.

I have always believed family plays an important role in shaping one's life. The love and encouragement of my grandparents from childhood and the pride of being the grandson of two very accomplished educators, Dr. Ramesh Chandra Pal and Dr. Gopendra Lall Seal is something I have always cherished. The timely, calm career advice of my Kaku, Mr. Ashok Pal at a rebellious phase of my life and the inspiration and support from my Mashibhai, Dr. Kaveri Das and Meshobhai, Dr. Goutam Das have all helped me in more ways than I can articulate.

A thank you to your parents is never enough. However, I would like to take this opportunity to express my appreciation to my mother, Mrs. Anuradha Pal, my father, Mr. Samir Pal and my grandmother, Mrs. Gouri Pal for all that they have done for me. I have been often insulated from all the hardships and sacrifices that they have undertaken to give me a comfortable life. Lastly, a big thank you to my wife, Mrs. Shreejata Maity for believing in me whenever I was low on confidence. The support, love and patience that she has bestowed upon me over the years is unparalleled. *"To our life together and beyond..."*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Ageing is a global phenomenon with the population of older adults of the world increasing at an alarming rate. According to a United Nation report, population ageing is unprecedented and pervasive with deep implications to every human life [2]. The report also states that "*The older population is growing faster than the total population in practically all regions of the world and the difference in growth rates is increasing*". The continual increase of the population [the statistical trend is presented in Chapter 2] of older adults all over world envisages the use of technology to promote better living standards and quality of life. Assisted living solutions mostly focus around the detection of falls which is a major risk for older adults. However the increasing need and scope of wider use technology have slowly found their applications in personal security, health and rehabilitation and even in bio-robotic systems. With the advent of novel communication technologies, increasing number of technologies are finding their application in prognosis of health conditions by monitoring one or several parameters in one's daily living.

Technology to monitor the parameters which affect the quality of life or lifestyle are being monitored using a number of different types of technologies like body-worn alert device, mobile based applications, intelligent clothing and passive infra-red sensor (PIRs) within assisted living environments. Different types of technologies are finding prominence and gaining traction for monitoring indoor as well as outdoor activities. Most assisted living environments or large scale observational studies have traditionally employed a network of sensors. With the huge reduction in the camera prices, the use of visual sensors are gaining prominence within assisted living solutions. Though some studies as mentioned in Chapter 2 have had visual-sensors as a part of their sensor network, images and videos were mostly used for post observation analysis.

Data acquired from a single sensor or a network of PIRs are often not contextual enough to deduce much information. For example, a simple motion detector can detect only motion within a room, however with the motion information of a video, we can detect motion, track an object, understand the direction of motion and also the intensity of motion. Visual sensor based technologies involve capturing of visual information, processing it and then analysing it. One of the fundamental advantages of a visual sensor-based system over a non visual sensor-based system, is that visual data obtained in contextually rich. However, it is this fundamental advantage of a video-based system that poses the biggest challenge against developing such a system as along with being contextual is also extremely sensitive as shown in Figure 1.1 (adapted from [3]).

The potential challenges while designing such a system as mentioned in [3], are

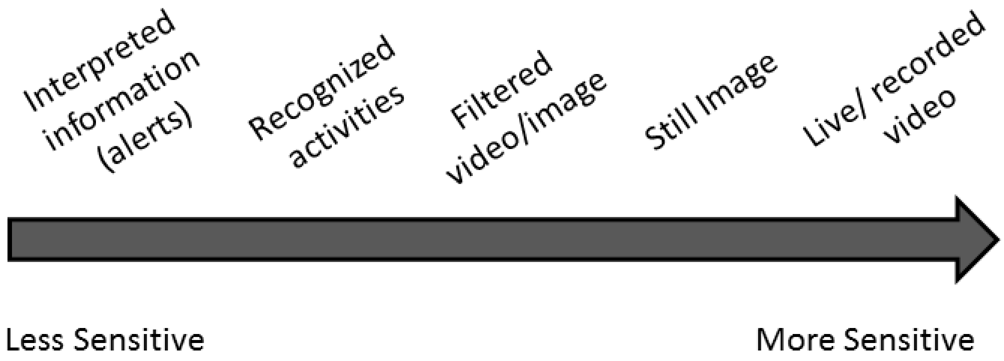


Figure 1.1: Sensitivity of Information

technical, acceptability and integration.

Technical Challenges - The technical challenges involve selection of a correct camera with good resolution. Along with the positioning of the camera, it is also important to note that visual data need a large bandwidth for transmission. If data is sent over a network to a remotely located server for a third party viewing and analysis, there should be stringent network protocols in place to prevent unwanted interception. There are smart cameras in which the video can be filtered and augmented symbols of the target or silhouette of the region of interest could be transmitted. Data over a network should also be securely and privately transmitted to avoid issues with data-protection. Also, in a typical home environment, there is a possibility of occlusions due to furnitures and other objects. A multi camera view of the room can possibly reduce the effects of occlusions but for a single camera set-up, the problem of occlusion becomes exponentially more. In daily living, it should be also considered that all the postures and locations within a room will not be always facing the camera, which in turn poses the challenge of self occlusion. Monitoring movement of hands to recognize activities or reconstruction of the scene becomes extremely complex due to self occlusion.

Acceptability Challenge - User acceptance is the most critical part specially when the user and the movements would be continuously video-captured for analysis. While a filtered image is relatively less sensitive, there is a need of more rigorous qualitative assessment regarding the acceptability and data protection protocols to understand the sensitivity of information regarding monitoring technologies. Ensuring privacy and protection of personal data without hindering the overall independence of an individual forms an extremely important bench-mark for the design and implementation of any monitoring technology. A balance between the context of the visual data and the sensitivity of the interpreted visual information is needed. The data need to be protected and kept confidential by several privacy filters and multi-level protection protocols.

Integration Challenge - The infrastructure and the ambient conditions of a dwelling form an important priori knowledge to a video-based monitoring system. The arrangement of the rooms, the lighting conditions, the colour of the walls, positioning of furnitures are important bits of information that needs to be addressed while selecting and designing the video-based system. Apart from the visual features, other relevant parameters like temperature, smoke within the dwelling should also be monitored for health and safety. An intelligent monitoring system should be able to effectively communicate with other devices and other data sources for monitoring not only the user but also other relevant parameters.

1.1 The Thesis

The research undertaken for this thesis forms a part of a multi-disciplinary network within The University of Sheffield. The motivation behind the network was to find out ways to promote independence among the older adult population within an assisted living environment. The outcomes of the network was aimed to identify the some of the translational requirements of using novel technologies for older adults.

1.1.1 PIPIN - multi-disciplinary research network

PIPIN (Promoting Independence through Personalized Interactive Technologies) is a multi-disciplinary network within The University of Sheffield focussing on the examination of existing and future technical capacity that can be used to benefit older people with long term conditions. The three projects funded within the network are

- Project 1 involved the technical translational of requirements of health and social care professionals for lifestyle/behavioural monitoring using visual sensors taking into account the concerns of end-users.
- Project 2 involved the development of a personalized adaptive listening system primarily for dysarthric speakers.
- Project 3 investigated the gaps in how specialized assistive technologies are appropriated, adapted and rejected in everyday settings.

This thesis is the outcome of the Project 1 whose main focus was to collect and analyse activity levels with visual sensors which can contribute to the overall activity profile of an individual. Traditionally these measurements were done using a network of low-cost sensors like motion-detectors, bed occupancy sensors and wearable sensors. Since visual sensors for automatic analysis of visual data had not been used before, a part of the research was also to investigate the context and acceptability of visual sensors for lifestyle monitoring through interactions with end-users.

[The author would like to note that the research presented in this thesis has been exclusively the author's work and research from other projects have not been included in this thesis.]

1.1.2 Aim and Objectives of this Thesis

Since the challenges to having a visual-sensor are multi-disciplinary, the research does not solely aim to solve the existing activity recognition or action recognition problem as is understood within the computer vision community. The aim of this thesis is to propose an effective metric for detecting the amount of activity carried out by an individual in an assisted living environment using visual sensors, taking into account the context of acceptability and privacy through an inter-disciplinary study. The aim of this thesis is achieved by the following objectives:

- to understand the context and acceptability of vision-based solutions for assisted living,
- to propose an effective metric (activity level) for the amount of activity carried out and a representative evaluation test-bench to test vision-based solutions,
- to evaluate the efficacy of the proposed metric using different visual features using a discriminative model of features and classifiers.

ASSUMPTION

Since the network of which this research is a part of focussed on promoting independence, for this research it has been assumed that the assisted living environment would be single occupancy and that the only moving subject within a scene would be the individual to be monitored. Also, it was assumed that the research would focus on normal daily living of individuals and not focus on any specific health conditions.

1.1.3 Author Contributions

The main contribution of this thesis is to propose a visual sensor-based activity level estimator for assisted living in the context of the amount of movement undertaken by an individual to carry out activities of daily living through an interdisciplinary study. The thesis is an outcome of one of the first studies that investigates the use of visual sensors for activity level estimation for monitoring of older adults within assisted living. The novel outcomes of this thesis -

1. Qualitative analysis based on data collected from prospective end-users of a video-based monitoring system of older adults. [Chapter 3]

2. Definitions of activity levels estimated with a visual sensor and a novel annotated bench mark dataset for activity level estimation [Chapter 4]

3. A novel activity level estimation system methodology under single camera setup and dual camera setup using 3 different features -

    - motion-HOG feature for activity level estimation [Chapter 5]

- novel motion-moment feature for activity level estimation [Chapter 5]
- novel localized phase correlation method for activity level estimation [Chapter 6]

1.1.4 Author Publications

Some of the contributions and research of this thesis have been disseminated in the following manner.

1.1.4.1 In Peer-reviewed Proceedings

C1. S.Pal and C.Abhayaratne, "Video-based Activity Level Recognition for Assisted Living Using Motion Features", in *Proceedings of ACM International Conference on Distributed Smart Cameras* (ICDSC), 2015, pp. 62-67.
C2. S.Pal, T.Feng and C.Abhayaratne, "Real-time Recognition of Activity Levels for Ambient Assisted Living", in *Proceedings of IEEE International Conference on Consumer Electronics* (ICCE-Berlin), 2015, pp. 485-488.
C3. S.Pal and C.Abhayaratne, "Phase Feature-based Activity Level Estimation for Assisted Living", in *Proceedings of IET International Conference on Technologies for Active and Assisted Living* (TechAAL), 2016.

1.1.4.2 Poster presentations

Along with the proceedings, portions of the research in this thesis has been disseminated as poster presentations. The posters and the accompanying abstracts were peer-reviewed and accepted by established and distinguished academics who might not have been domain experts. The posters were presented in the following events:

P1. S.Pal, C.Abhayaratne, M.Hawley, "Video-based Monitoring of Activity Levels for Assisted Living" in Proceeding of Dem@Care Summer School on Ambient Assisted Living (DemAAL) [2013].

P2. S.Pal, C.Abhayaratne, M.Hawley, "Video-based Activity Level Monitoring for Assisted Living". Presented at the Regional Event of Royal Academy of Engineering, UK [2015] [**Adjudged first-runners up prize**].

P3. S.Pal and C.Abhayaratne, "Detection of Activity Levels for Monitoring of Daily Activities of Older Adults in Assisted Living" in Proceedings of Workshop of Human Motion Analysis for Healthcare Applications, IET, 2016.

1.1.4.3 Journals under preparation

J1. Visual Sensor-based Monitoring Framework for Assisted Living - An Eidetic Log of One's Lifestyle. [IEEE Journal of Biomedical and Health Informatics]

J2. Activity Level Estimation for Assisted Living using Motion Features in the Pixel Domain. [IEEE Transactions on Consumer Electronics]

J3. Localized Phase-based Activity Level Estimation for Assisted Living. [IEEE Transactions on Systems, Man, and Cybernetics]

1.1.5 Organization of the Thesis

Chapter 2 details the statistical trends in the ageing population of the world to justify the need for technological solutions to provide care and support. Following the statistical trends, commonly used monitoring systems are discussed followed by video-based monitoring systems. In this section a few popular activity recognition algorithms are discussed. Though the thesis is not about activity recognition, the work done in the literature are mostly focussed to solve the activity recognition/action recognition problem without taking into account the context of application.

Chapter 3 is about the qualitative study of a video-based monitoring system to meet the first objective. The need for focus groups as a mode to understand the context and use of a video-based system is detailed. The structure of discussion in the focus groups among older adults of the community is listed along with the outcomes of the discussion. Along with the analysis of the focus group discussion, 4As model is proposed for visual-sensor based monitoring is introduced.

Chapter 4 defines the concept of activity levels and how the definitions have been formalized through a visual perception test to meet the second objective. In this chapter, a short review of existing datasets is presented highlighting the need for the novel dataset is also presented. The Sheffield Activities of Daily Living dataset is proposed along with its technical specifications.

Chapter 5 A novel activity level estimation system is introduced. Initially simple motion features are use to estimate activity levels which forms the baseline results. The results are improved by exploring different motion-based features to detect activity levels in the pixel domain. Keeping the classification parameters constant, two different features are compared under two different camera setups. The proposed motion-moment perform the best in this comparative study.

Chapter 6 contains estimation of activity levels in the frequency domain. Inspired by the psychophysical aspects of motion perception and the information offered by phase images of a Fourier Transform, a novel localized phase correlation method is proposed to model activity level. The results are compared to the pixel domain results to meet the second part of the third objective. Using the properties of the Fourier space, phase information is used as a feature to estimate activity levels.

Chapter 7 forms the conclusions to this thesis and lists possible future directions of this research.

Chapter 2

Lifestyle Monitoring for Assisted Living - A Background Study

Lifestyle Monitoring technologies are information and communication technologies which monitor individuals in their daily living to identify early signs of health deteriorations. With the emergence and easy availability of communication technologies like mobile phones, many mobile application led solutions have also evolved to monitor day to day activities of individuals. The rest of the chapter is as follows, section 2.1 details the motivation behind having lifestyle monitoring technologies for older adults. Section 2.2 lists some of the interpretations of what a lifestyle monitoring technology is in the context of assisted living for older adults, the need for lifestyle monitoring and also list the different types of non-visual sensors that are used of user-trials and commercial installations. Non-visual network of sensors are often obtrusive and complicated and the information deduced from sensor firings are not always accurate. Visual data on the

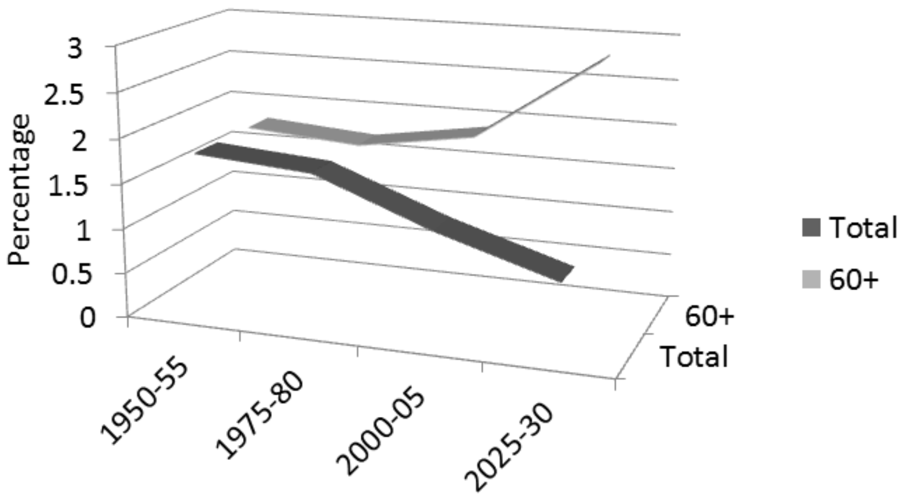


Figure 2.1: Average annual growth rate of total population and population aged 60 or over: world, 1950-2050

other hand provides more contextual and precise information. Section 2.3 details the use of visual sensors for assisted living and how different approaches have been adopted to address different requirements of such a technology. Finally, in section 2.4, a summary of the chapter and the research focus for this thesis is detailed.

2.1 Ageing Population

As shown in Figure 2.1, the growth rate of population of the world has decreased from 1.8% to 1.2% between the years 1950 and 2000 and is projected to decrease further to 0.8% in 2030. However, the change in the older adult population is expected to rise from 1.9% to 2.8% between 2000 and 2030. The demographic changes of a few countries as shown in Figure 2.2 show that almost all of them would experience a sharp rise in the percentage in the older adult population [4].

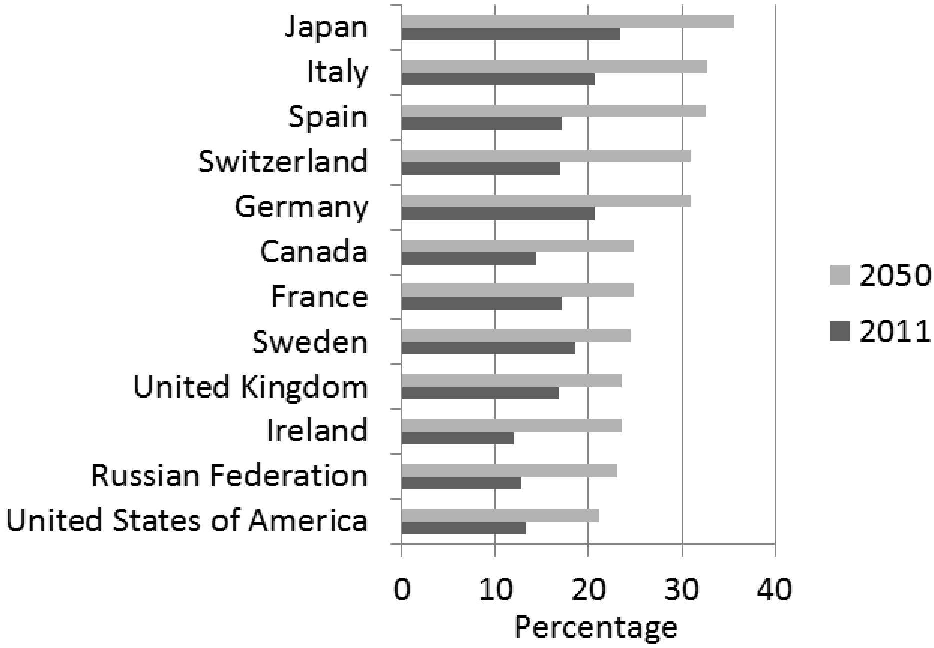


Figure 2.2: Percentage of population aged above 65 years

While Japan is supposed to have the highest percentage of older adult population with 35.6%; in countries like Ireland, Switzerland and Spain, the percentage of older adults would double by 2050. In other European countries like Germany, France and Sweden there would be an increase between 7% to 12%. Even within the United Kingdom, the population aged 65 years and above is projected to rise by 6.8%. As shown in Figure 2.3, the number of people between the age 60 to 75 years would rise by 2 million and in the above 75 years category, the rise would be nearly 4 million [5]. Since the overall life expectancy of human beings has increased owing to the advancement of medical sciences and better health care services, there is a continuous growing need to provide better social and healthcare services to the ageing population.

Increase in the population of older adults envisages an increase of dependants on the family and the Government. Based on [6], as many as 35% of older

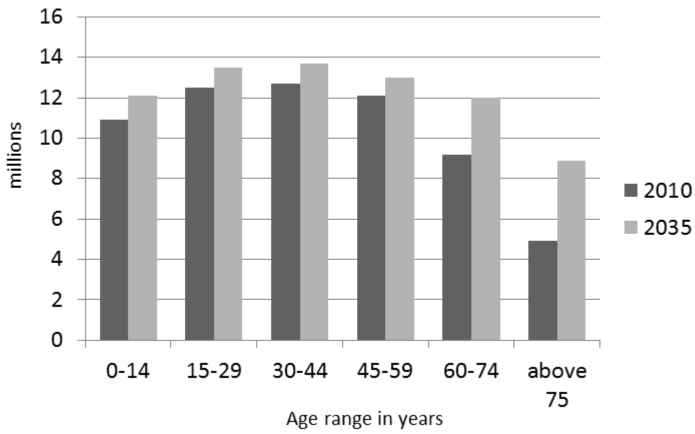


Figure 2.3: Projected population by age, United Kingdom, 2010-2035

people who live in a care home could be supported to live at home or in extra care housing schemes through the use of technology. Along with the current care models, tele-care systems, and the rise in health-care costs; the society would need more robust and cost-effective technology dependent monitoring solutions that would help monitoring the well-being of the older adult population. Emerging monitoring technologies should be low-cost, simple, unobtrusive, easy to interact and would be able to promote an independent lifestyle of the ageing population.

2.2 Lifestyle Monitoring

With the population ageing at a rapid pace, the need for assisted living environments and lifestyle monitoring technologies is increasing. Assisted living environments are spaces enabled with information and communication technologies that would assist in one's independent day to day living. The emergence of modern information and communication technologies have led to an increased number of research and commercial implementations of different types of technologies to aid the older adult population to lead a more healthy lifestyle and also to predict an impending life condition.

2.2.1 Definition

Modern telecare and telehealth solutions focus on using technologies to aid in an healthier lifestyle and prognosis of a health condition through indirect measure of how active an individual has been in daily living. One of the earliest studies of behavioural monitoring of individuals done by Celler *et al.* [7] found out that the health status of an individual can be determined by monitoring a number of simple parameters (mobility, sleep patterns, utilisation of cooking, washing and toilet facilities) which are representative of the interaction of the individual with his environment. There is no universal definition adopted for the term lifestyle monitoring for older adults. As mentioned in [8], lifestyle monitoring for older adults can be termed as one of the many elements of telecare research which aims to detect changes in the activity profiles of an individual through one or many measures in order to highlight an individual's health status or care status. Lifestyle monitoring within assisted living stems from the hypothesis that there is a correlation between health and activity [9].

2.2.2 The relationship between Activity and Health

The relationship between physical activity and health status is well recognized, investigated and recommended by The World Health Organization (WHO) [10]. The inter-relationship between health and activity has been a topic of scientific enquiry for many decades leading us to broadly conclude that a sedentary living can have its implications of the physical and mental health of any individual [11]. Though there is no quantitative measure to ascertain the correlation between health and activity, it is widely accepted that physical activity contributes to healthy living among all age groups.

Increased mobility or physical activity improves one's psychological well-being and quality of life [12]. A decrease of Physical Activity Level or physical functionality can reduce the functional dependence and increase the risks of disability with age [13] [14]. Activity Intensity or the energy cost of physical activities have long been manually measured by a physiological measure named metabolic equivalent task (MET). This measure is a ratio of the metabolic rate for a particular activity to the set of reference values defined for a group of activities. For older adults, the amount of daily activity undertaken also becomes indicative or precursive knowledge of one's failing health condition [8] [15].

2.2.3 Activities of Daily Living

Activities of Daily Living are measures by which one's functional ability to live independently is estimated. Way back in $1960s$, Katz had proposed the Index of Independence based on the activities of daily living (ADLs) [16]. The activities of daily living are parameters to represent the functional ability of an individual to carry out an activity. Activities of Daily Living are categorized into BADLs (Basic Activities of Daily Living) like bathing, dressing *etc.* and IADLs (Instrumental Activities of Daily Living) like housework, shopping *etc.* The BADLs are about fundamental functions of one's living whereas the IADLs tell us more about independent survival [17]. According to Gil *et al.* in [18] there are two basic approaches to measure activities of daily living of older adults -

- Monitoring the ADLs by looking at specific activities.
- Measuring the activity or busyness of an individual.

Measuring activities of daily living has often being criticized as asking an individual to carry out a list of activities which will be representative of healthy living fundamentally violates the concept of independent living. Also, different individuals would have different activities of daily living and hence would carry out different activities differently as demonstrated in a study done by Dodge *et al.* [19], where variability in walking speeds and trajectories were correlated with mild cognitive impairment. Hence, instead of asking an individual to carry out a fixed set of activities, having a proxy/indirect measure to the amount of activity is more acceptable for monitoring lifestyle of an individual.

2.2.4 Non-visual Sensor-based Technologies

Lifestyle monitoring technologies are information and communication technologies which help monitor the lifestyle of an individual in daily living by monitoring one or more parameters through one or more sensors. The parameters of monitoring lifestyle range from observing activities of daily living to monitoring various ambient parameters. Most commercially available sensor-based technologies or those developed for user trails involve a large number of embedded sensors like motion detectors, accelerometers *etc.* to monitor a single activity of daily

Table 2.1: Commonly used sensors and their domains of application

Application Domains	Sensors
Activity monitoring	Motion detector, Door open/close, Electrical appliance, Accelerometer
Safety and Security	Motion detector, Door open/close, Electrical appliance, Temperature/Smoke sensor, Accelerometer
Fall detection	Motion detector, Microphone, Accelerometer

living [20]. Sensors used for the lifestyle monitoring can be broadly divided into wearable sensors and non-wearable sensors. While the wearable sensors like accelerometer, smart clothing (gloves, shoes *etc.*)are used for monitoring daily activity patterns as well as collect physiological information; the non-wearable sensors like motion sensors, temperature sensors *etc.* are used to monitor the ambient condition for security and safety. The application domains of a monitoring technology can be broadly divided into activity monitoring, personal safety and security and fall detection. Some of the most commonly used sensors in these domains are listed in Table 2.1 [adapted from [3]].

2.3 Visual sensor-based Lifestyle Monitoring Technologies

In recent years, there has been a great interest to explore the applications of video based technologies in health research and assisted living. Video cameras have been used in some trails but most of them were used for post-observation manual analysis of the recording. For example in the *GEROME* project [21], video cameras were used along with audio sensors for enhancing independence of the elderly at home by ensuring autonomy, comfort of life, security, monitoring and assistance to place of residence. Also in the *HERMES* project [22], video cameras and audio sensors were used to provide cognitive care for the elderly. However in both these studies, visual data was used more for manual post behavioural analysis. Visual sensors have their advantages over non-visual sensors as they are able to capture more than 1 event at a time and also the data obtained is data rich [3] [1]. Visual data is also contextually rich and provides more information than non-visual sensor firings.

2.3.1 Alert Systems

The most common application of a video based technology for assisted living is fall detection. A majority of the literature focusses on different fall detection techniques using the bound-box approach, motion analysis, aspect ratio measurement *etc.* ([23], [24], [25]). Some of the other works include Chen *et al.* in [26], where the authors used 23 cameras and computer vision algorithms to detect elopements among people with dementia. Similarly in [27], a two way interactive video technology is used to monitor patients with mid dementia.

2.3.2 Activity Recognition

The term activity does not have a universal definition. Different disciplines tend to define activity in different ways. While in the rehabilitation field, it is defined as activities necessary for day to day living; in gerontology, even social participation is categorized as activities along with physical movement. The term activity is differently defined within computer vision literature and health service literature.

Within the computer vision community, activity recognition is a broad research topic. Activity recognition in context of assisted living is not as widely reported as generic activity recognition or action recognition. Some of earliest view-point dependent solutions used the motion history images as proposed by Davis and Bobick in [28], to model different actions in indoor settings. For example, Albu *et al.* in [29] proposed the Volumetric Motion History Image (VMHI) to analyse irregularities in human actions. The experiments suggest that along with the VMHI information, additional measures like speed might be needed for a more robust detection of abnormal behaviours. Weinland *et al.* introduced Motion History Volumes (MHV) as a free-viewpoint representation for human actions in the case of multiple calibrated, and background-subtracted, video cameras in [30]. Another approach involved having local descriptors based on spatio-temporal interest point like SIFT, SURF, HOG are often used to recognize actions. Computing these local features and then modelling it over time are resource intensive tasks and not often not discriminatory enough to model human motion [31]. Moreover, these approaches have seldom been tested in context of a home monitoring system or assisted living.

Within health service, an activity refers to more of activities of daily living and it is measured using various parameters whereas within computer vision activity is made up of one or more component actions with a high level semantic representation [32] [33]. As mentioned in [3], activity detectors can often be made up of stochastic context free parser composed of action detectors as sub-components.

Activity recognition is heavily dependent on the type of features one uses for the classification and the type of dataset the algorithm is used on. With the growing interest to use visual systems for home monitoring systems, Messing *et al.* in [34], introduced the velocity histories of tracked key points to model the activites of daily living. In [35], Cheng *et al.* have also used the local HOG features over time to model different actions. Along with the local features, Wang *et al.* in [36] have used depth information to model different actions into an ensemble of *actionlet* to label each activity. All these approaches have been evaluated on their own created datasets where the subjects in the video are facing the camera.

2.3.3 Other Parameters

One of the main problems of activity recognition in a video-based indoor monitoring system is occlusion. Occlusion in complex scenes can be detected by the segmentation based approach as in [37], where the video sequences are broken down into layers to represent the depth. A multiple level MHI (MMHI) was formulated by Valstar *et al.* in [38] to overcome the problem of self occlusion by taking the MHI at multiple time intervals. In [39], Stein explores the various occluding detection techniques such as the contour-focussed technique, the T-junction detection using image features like edge information and motion information into account. The main cause of occlusions in an indoor setting is due to furnitures or due to self occlusion. To completely remove the noise due to occlusion from an indoor scene is an extremely challenging task. Also, selection and positioning of the camera becomes important when we want to have minimal noise due to occlusion in the video sequences. Some of the other approaches used in video based assisted living are thermal imaging technologies for physiological information monitoring [40], action recognition [41], gait analysis [42] *etc.* More recently, Edlib *et al.* in [43], used a network of low resolution visual sensors to detect for visitors using silhouettes and the Markov model.

2.4 Discussions

This chapter focusses on the background of this thesis detailing the motivation of this research, the current lifestyle monitoring technologies and the use of visual sensors in addressing some of the challenges in isolation. In this section, a summary of the chapter is presented followed by identification of the requirement of a lifestyle monitoring technology and concluded with the research focus of this thesis.

2.4.1 Summary of the Chapter

As mentioned in section 2.1, the motivation behind the research is the rapid increase of the older adult population. The number of people to be cared for would soon surpass the number of people who can provide care which emphasizes the need for having information and communication technologies to promote independent living. Most telecare technologies, alert systems or anomaly detection technologies generate alert only after an event has happened. Most lifestyle monitoring systems used for user-trials are PIR-based or wearable technologies. However, as mentioned in [3] and [44] PIR sensors are often obtrusive, complicated, unreliable, have a limited range and offer limited information; whereas, visual sensors provide more contextual and precise information. Visual sensors have been used in various monitoring technologies but they have been used mainly of observational purposes and post-analysis purposes.

2.4.2 Requirements of a lifestyle monitoring system

Under the assumption that a decline in daily activities is an early indicator for failing health, the next generation of lifestyle monitoring technologies focusses on the prognosis rather than diagnosis of a medical condition. Activity measures are often indirect measures of the abilities of an individual to carry out an activity. These are often collected by self-reporting of the ability to carry out activities of daily living or health-care professional actively or passively observing the performance of activities [45]. Lifestyle monitoring technologies are technologies which can directly monitor an individual in daily living; thereby more emphasis can be given on the behavioural patterns or correlation of activities to declining health. So the requirement of lifestyle monitoring technologies, is having a measure of activity which would help in creating an activity profile for an user. Identification of activity for building activity profiles do work but in [9], the author argues that end users have expressed displeasure in carrying out a list of activities to fulfil the goal of healthy living. The end users would much rather live independently and go about their normal routine and have a proxy measure to determine if they have done enough activity.

2.4.3 Research Focus for the Thesis

The challenges of designing and evaluating a lifestyle monitoring system are multidisciplinary. In this thesis, the focus has been narrowed down to establishing and estimating a proxy measure for the activity profiles of an individual from visual data. The research focus for this thesis begins with identifying the requirement of a lifestyle monitoring system. The qualitative studies exploring the use of visual sensors for monitoring have focussed more on observation of activities rather automatic interpretation of an activity measure. Hence, the acceptability of using visual sensors in a lifestyle monitoring system is investigated through focus groups as presented in **Chapter 3**.

One of the requirements of a lifestyle monitoring technology is having a measure of activity which observed over a longer period of time could potentially help in generating activity profiles. Hence, a higher level of abstracted information about activities of daily living is proposed and evaluated through a visual perception test. Also, the lack of datasets for activity level estimation as mentioned in section 4.2, prompted the creation of a dataset with the ground-truth labelled with activity levels. The definitions of activity levels for visual proxy measure named as activity level are proposed in **Chapter 4**.

Activity recognition as understood within computer vision is recognizing of activities for various applications. Activity recognition in the context of assisted living or behavioural monitoring often focusses more on the recognition of activity and less on what is actually needed for a monitoring technology. Most activity recognition algorithms detail about the motivation and the need for assisted living but often fail to address the specific requirements of a monitoring technology. In

Chapter 5 and **Chapter 6**, different image features in the pixel domain and frequency domain are explored to estimate activity levels.

Activity Level as a measure of busyness of an individual estimated with visual sensors was first proposed in [46]. It is only recently that the SPHERE project [47] focusses on building an intelligent sensor platform for healthcare in a residential environment where the aim is to have a multi-modal solution using traditional sensors and video cameras and some of the visual data obtained through this platform is automatically analysed and processed for activity recognition tasks. It must be noted that the activity recognition task in this project as well is done by not reconginzing activities but by estimating activity intensities [48] based on energy expenditure and classifying them into 3 groups namely light, moderate and vigorous. Activity level estimation and activity recognition are two fundamentally different problems as one aims to find out the amount of activity while the other aims to recognize an activity.

Chapter 3

Visual Sensor-based Monitoring for Assisted Living : A Qualitative Study

Video-based lifestyle monitoring for assisted living is an emerging domain and it is very important to understand the context and implications of using cameras for monitoring of daily activities. As mentioned in Chapter 2, the activities of daily living are mainly measured through self-reporting. In the study done on patients in [49], the author concluded that though the inter-observer reliability of activities of daily living is excellent, patients often tend to over estimate their ability to carry out activities of daily living. Monitoring systems for older adults address this challenge by employing a network of non-visual sensors to unobtrusively decipher the actual of amount of activity done by an individual. Visual sensors for monitoring systems are slowly gaining prominence but mostly for post analysis for behavioural modelling. In this chapter, the context and acceptability of visual sensors for monitoring systems have been explored through focus groups among older adults within the community. Based on the outcomes of the discussions among the participants, a model for video-based assisted living system is proposed.

3.1 Introduction

Behaviour modelling using manual annotated video recordings is slowly gaining prominence for diagnosis and prognosis of long term conditions. With the continual reduction of camera prices and the evolution of computer vision techniques, camera-based solutions are becoming more and more low cost, simple and unobtrusive alternative for sensors. Visual data is contextually rich and often simple features from video data have lots of information about the scene embedded in them. As shown in Figure 1.1, the sensitivity of information acquired by a visual sensor is much more than that acquired by a traditional PIR sensor. Since visual data is extremely intrusive and can lead to privacy concerns, Focus groups were organized among older adults within the community who discussed about the possibilities and impacts of having a camera for monitoring at home. The privacy concerns of older adults are contextual, personalized,and influenced by psychosocial motivations of later life [50]. Hence, the degree of abstraction in extracting, processing and analysing data for an assisted living technology is of utmost importance.

In this chapter, a novel video-based monitoring model is introduced where the analysis of the video-recordings would be done automatically using different computer vision techniques. Focus group discussions among older adults from within the community were conducted to explore the context and acceptability of a visual-sensor based monitoring system. The novel outcomes of this study are -

- a qualitative analysis based on data collected from prospective end-users,

Along with the analysis of the focus group discussions, a video-based lifestyle monitoring system architecture is introduced. The architecture has a high level representation as well as a more granular representation. The rest of the chapter is arranged as follows: Section 3.2 details the background to user-centered design and other studies which have explored the visual sensors for monitoring systems. Section 3.3 discusses the methodology adopted to understand the context and acceptability along with the structure of the focus groups. The analysis of the focus group discussions is presented in section 3.4 followed by the multi-level model in section 3.5. Finally, a summary of chapter is presented in section 3.6.

3.2 Background

Telecare services focus on alarms and monitoring of activities. The fundamental essence of a monitoring system is to provide a safe and secure environment, thereby promoting a healthy, independent and non-intrusive living. Traditional system development life cycle (as shown in Figure. 3.1) does not involve users until the testing phase.

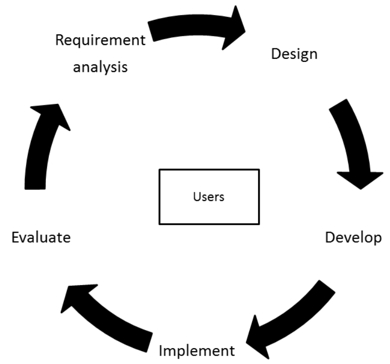


Figure 3.1: System Development Life Cycle

However, for a monitoring technology, acceptance and usability of a technology are the two most important questions that need to be addressed at the very onset. To understand the acceptance of a video-based monitoring system and the privacy concerns of such a system, an user-centred design approach should be adopted. In a recent study done by Bentley *et al.* [51], the authors suggest that inspite of the advances in technology in recent years, there are still considerable design related issues in current telecare systems. In other study done, the authors note that a strong involvement of end-users is needed to address usability and accessibility of monitoring technologies, thereby stressing on a more user-centric approach to design and development of technology [52].

3.2.1 User-Centred Design approach

The four core activities on designing usability according to the International Organization for Standardization [ISO 13407] on usability [53], (as shown in Figure 3.2) are -

- **understand and specify the context of use** - identify the need and

Table 3.1: ISO-13407 activities and system development life cycle

Activity	System Development Life Cycle
understand and specify the context of use	Requirement analysis
specify the user and organisational requirements	Requirement analysis
produce design solutions	Design and Implementation
evaluate designs against requirements	Testing and Evaluation

context of relevance of the system.

- **specify the user and organisational requirements** - the actual requirements and the success criteria of the system

- **produce design solutions** - create innovation solutions in accordance to the requirements

- **evaluate designs against requirements** - testing and evaluating the system.

Each of the above activities can be related to each step of the system development life cycle (Figure 3.1) and are listed in table 3.1.

User centred design refers to an iterative system development process (as shown in Figure 3.1) in which the end-users contribute to the design and evaluation. According to Norman [54] "*user-centered design emphasizes that the purpose of the system is to serve the user, not to use a specific technology, not to be an elegant piece of programming. The needs of the users should dominate the design of the interface, and the needs of the interface should dominate the design of the rest of the system*".

Three of the most popular user centred design methods as listed in Table 3.2 for requirement elicitation are focus groups, questionnaires and interviews. The

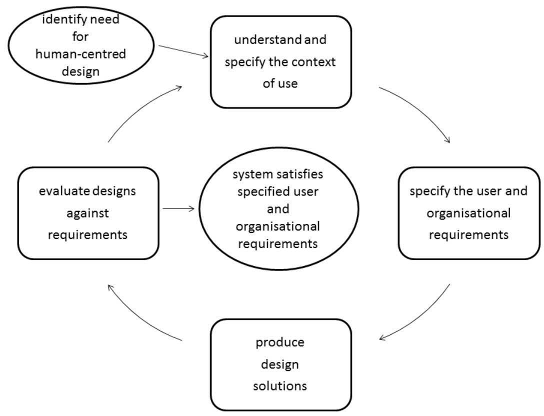


Figure 3.2: Activities of user-centered design

Table 3.2: Comparison UCD methods

Methods	**Output**	**Sample-size**	**Cost**
Focus Groups	Non-statistical	Low	Low
Questionnaires	Statistical	High	Low
Interviews	Non-statistical	Low	High

cost and the sample size of each of this methods are listed in Table 3.2[adapted from [55]]. Using visual sensors in a behavioural monitoring system for automated analysis of visual data is a concept that is slowly gaining prominence. Though visual sensors are increasingly finding their applications in surveillance, the use visual sensors for monitoring system is relatively new specically when the system automatically analyzes the visual data into information. Hence to investigate the context and acceptability of having visual sensors, focus groups is the most preferred methodology. For example in [56], the authors investigated the different features of a smart home through focus groups. Also, in [57], in a comparative study between interviews and focus groups for ensuring acceptability for behavioural and social interventions, the authors note that the advantage of focus groups is that it enables the participants in collective brainstorming about a topic.

3.2.2 Other Studies exploring Visual Sensors for Monitoring

Ensuring privacy and protection of personal data forms an extremely important bench-mark for the design and implementation of any monitoring technology. User acceptance is the most critical part specially when the user and the movements would be continuously video-recorded for analysis. Londei *et al.* [58] in their study have found out that nearly more than 80% of the participants in their interviews had a positive response to an intelligent video-based monitoring system to detect falls though the participants did make a distinction between image analysis executed by a computer and the images seen by authorized personnels. Boise *et al.* in their study [59] found out that over 72% of participants reported acceptance of in-home computer-based monitoring and willingness to have data shared with their doctor or family members. Along with their willingness to a video monitoring system in [59], around 60% of the participants have expressed concerns related to privacy or security. Though these studies have investigated the use of visual sensors or cameras for monitoring, they have mostly been in context of a specific health condition or for an alert system like fall detection. In this work, the use of visual sensors for a monitoring system is investigated for a more general purpose use for behavioural or lifestyle monitoring.

3.3 Methodology

Video-based monitoring is an emerging domain within assisted living research. The primary aim of the focus group is to understand the attitude of older adults and specify the use and context of a video-based monitoring technology. Though most of the participants had an idea about technology and how it can help us, most of them were unaware of emerging computer vision applications and how visual data can be automatically analysed to obtain information. To organize the discussions, ethical approval was sought from the University and participants were recruited from local voluntary organizations. Finally, the discussions were structured in such a way that initially they were given an overview of the monitoring technologies and how visual-data based applications are being increasingly used commercially in the domain of surveillance, entertainment industry *etc.*

3.3.1 Focus Group as a Qualitative Tool

For a new and emerging concept like video-based monitoring, focus groups are the most relevant. Focus groups are more like a general discussion where participants express their opinions spontaneously on a particular topic [60]. Focus groups are extremely useful to gather opinions and perceptions of people regarding a particular topic. Focus groups have been used to explore the effects of films and television programmes within communication studies [61]. They are also used to study health behaviours and to explore attitudes and needs of employees within corporate houses. Two of the most obvious limitations of a focus group are having a dominant participant and unconsciously introducing a moderator's bias [62]. However, to understand perspectives and attitudes focus groups are always the preferred choice because unlike interviews or questionnaires, in focus group it is the participants who drive the discussion rather than the interviewer.

3.3.2 Ethical Clearance

Ethical clearance from The University of Sheffield was obtained prior to the study. The clearance was granted by the review committee (comprised of the reviewers from The Faculty of Engineering and School of Health and Related Research) only after carefully examining the application form, the topic guides and the consent form. The Ethical Clearance application (along with the topic guide and information sheet) can be found in Appendix E.

3.3.3 Recruit Participants

The participants of the focus groups were recruited from among the members of the community. Anyone who potentially fulfilled the recruitment criteria were invited on a first-cum-first basis. No specific expertise, technical knowledge or disclosure of any medical condition was mandatory. In fact the participants were encouraged to participate and engage in the discussions based on their spontaneous feelings of the proposed technology.

3.3.3.1 Inclusion Criteria

Any individual who are above the age of 70 years and are healthy could participate in the discussions. Participants should enjoy living independently and would be physically mobile in order to commute to the location of where the focus group would be held. Participants must be living in the community and were expected to have adequate English language skills in order to understand and follow the instructions. Also the participants had to give consent to being audio-recorded as a part of the focus group participation.

3.3.3.2 Exclusion Criteria

A potential participant who has any severe medical condition that may affect the understanding and participation process was excluded. Participants with severe sensory deficits and visual impairments that cannot be corrected with glasses or contact lenses to within normal or near normal limits were also excluded from participation.

3.3.4 Stages of Focus Group Discussions

The different stages of discussion were as follows:

Introduction - this stage was used to introduce the participants to each other and also lay down the ground rules of the session. The ice-breaker question clearly revealed that most of the participants were accustomed to technology in some form or the other; the most common being a mobile phone.

Evolution of monitoring technologies - this stage was used to inform the participants about the current sensor-based monitoring system. The participants were also introduced to the idea of how of computer vision algorithms can be used for a variety of applications by showing them demonstration videos. The discussion pointers for this stage were:

- How do you feel about such a technology?
- What do you think about camera-based solutions?
- What are your thoughts about a monitoring technology?
- Do you foresee a video-based solution for a monitoring technology that you would like to use for yourself or a family member?

Video-based monitoring system - the participants were shown a video of a prototype software which indicated how active a person is to initiate discussion about their acceptability and concerns about a video-based monitoring system. The discussion pointers for this stage were:

- What sort of activities/events should the technology identify?
- Where and how would they like the information (video/health) be stored?
- To whom should the information be disclosed to? With whom should the control to access the information be with?

Concluding remarks - the participants commented on their experience of participating in such a discussion.

All the stages were aided with presentation slides or examples videos or demonstrations and concluded with discussions among the participants. The selected participants for the discussions were all retired personnels. The ratio of female participants to male participants was nearly equal to one. This helped in understanding the privacy concerns of a monitoring system across both the genders.

The discussions were audio-recorded and later transcribed for analysis with the consent of the participants. The initial analysis provides the overall attitude of the participants before seeing the prototype video and how it changed after seeing the prototype video. The presentation shown during the focus group discussions is presented in Appendix E.

3.4 Thematic Analysis

The data collected from the focus groups were audio-recorded which were later transcribed for analysis. The data was initially analysed and then further categorized into different themes. Thematic analysis as defined by Braun *et al.* in [63] is "*a method for identifying, analysing and reporting patterns within data*". For gaining acceptance of new ideas, it is very important to understand the existing attitudes of people [64]. The themes are further categorized as primary outcomes which involves the main points of discussion and the secondary outcomes which were discussed as complimentary points.

3.4.1 Initial Analysis

Within the thematic analysis framework, exploring for themes is mainly done using two approaches, one inductive reasoning and the other deductive reasoning. The initial results were encoded from primarily the second stage of the discussions to get a perspective about how the participants felt about monitoring technologies. As shown in Table 3.3, though the participants had an understanding of technology and its wide-spread applications in different domains in the modern world, there was still hesitations about how actually a camera can be used for monitoring without actually someone looking into their recorded videos. After seeing some examples of how computer vision techniques are being increasingly used in some surveillance and other creative industries applications, the participants slowly gained confidence about the proposed discussion topic. The participants started to have a better understanding of how a video can be automatically analysed to make out meaningful information without actually someone seeing the video and slowly started to reflect on the possibility of a video-based monitoring system which could be useful and helpful. Though there were still mixed reactions among the participants about video-based monitoring; almost all participants had examples from their own lives where they have faced multiple instances where there has been an emergency situation with one or more of their ageing relatives or themselves. All the participants agreed that if there was a monitoring system, then the medical services or family could have been alerted more promptly. Along with referring to these real-life experiences, the participants were eager to know about the possibilities and advantages of a video-based system. They were ready to push the boundaries of their individual privacy if a possibility of independent living in one's own dwelling exists; as one participants confidently claimed, "*We will get used to it*".

3.4.2 Primary Outcomes

The primary outcomes are the analysis results of the overall participant responses around which the maximum discussion took place. The commonalities among the responses of the participants are clustered into the 7 primary themes. Each of the

Table 3.3: Participants' responses before seeing the prototype video

Participants' responses	Perceptions Before	Perdeptions After
1. "That's a bit like a Wii" 2. "Burglary alarms"	Current sensor-based solutions	General awareness of technologies
1. "My brother lives in Germany and we got Skype up for my mother" 2. "Are those cameras in my block of flats reactive or are those storing videos"	Video-based solutions	
1. "Who will do the watching? That gets round to the privacy thing?" 2. "One might feel that it may invade your privacy. It is about striking a balance between privacy and safety"	Initial Hesitations	Attitude towards Video-based solutions
1."most of the surveillance at the moment, there is somebody sitting in front of the screen looking at it, whereas in this case, it the computer interpreting it", 2. "The physical concierge in my block of flats was replaced with cameras and I did not want it, but 6 months down the line, I don't have a problem with it"	Gaining Confidence	
1. "My wife collapsed with stroke a few years ago and lay there for like 12 hours before anyone found out 2. "I think I want to be independent and be on my own. So this would be perfect to have somebody to keep an eye on me."	For Self	Foreseeable possibilities
1."Well I can see how this would be fantastic for my parents. My mum had dementia and my father has a whole lot of physical disabilities." 2. "My mother-in-law spends 11 hours at night sometimes on her own, basically trapped in her bed with no surveillance other than if she could reach that button. And then this would would be reassuring for everybody."	For Family	

first 2 themes are broken into two sub-themes where the first sub-theme deals with the attitude of the participants towards a monitoring system while the second sub-theme focusses on the attitude of the participants towards having a video-based solution to the monitoring system. The remaining 5 themes mostly focus on the privacy concerns and their vision of a video-based monitoring system.

3.4.2.1 Theme 1 : Awareness and possibilities

The participants had an understanding of technology and its wide-spread use in the modern world. They were generally aware about how technology is being increasingly used in different aspects of one's daily living.

Sub-theme 1.1 : Technology in our lives

Almost all the participants had real life examples of how they have faced multiple instances where there has been an emergency situation with one or more of their ageing relatives or themselves and how it could have helped them if only there was a monitoring system that could have alerted the medical services or the family about the emergency. They were aware of sensor-led interactive technologies like 'Wii' or how motion-sensors are installed in houses for burglary alarms. Some of them were even aware of a few wearable technologies or calling buttons that older adults often use. However, the idea of the possible use of the video-based system for a monitoring technology was mostly limited to how camera-led technologies have revolutionized communication and the ease with which one can have video-conferencing calls using readily available free web-based technologies as noted by one of the participant "*My brother lives in Germany and we got Skype up for my mother*".

Sub-theme 1.2 : Cameras are being readily used

Along with the video conferencing technologies, the participants were aware of the some of the readily available technologies and the fact that CCTVs are commonly used for surveillance. The prevalent idea was a camera is there only for recording videos and someone would be physically watching them as it happens in most surveillance applications like a concierge of the building society or a shop. One of the participants inquisitively posed the question, "*Who will do the watching? That gets round to the privacy thing*"? regarding using a camera for monitoring. From such a viewpoint, the idea of having a camera within their homes and continuous monitoring of activities was not readily appreciated. Once the participants were confident that the camera is being essentially used to capture abstracted visual information of the scene for monitoring, the participants became comfortable with the idea with one of the participants claiming, "*I was a bit sceptical initially, but now I think that this is a good idea*". They were particularly interested if the technology could issue an alert during an emergency when the user or the monitored person would not be in a position to call for help. However, there were still concerns about how reliable the software is and whether it can be trusted completely for monitoring scenarios.

3.4.2.2 Theme 2 : Independence and safety

Participants expressed a strong desire for independent living and wanted to live in their own homes for as long as possible but they were concerned about their safety as well.

Sub-theme 2.1 : Independent and safe living

One of the participants claimed that one of the downside of a wearable technology is when one would have a visitor, the tendency is to take it off because "*she did not want to be seen to be dependent*". There was a particular dislike for wearable technologies as most of them pointed out that the fundamental disadvantage of having a wearable is often the user has forgotten to wear the technology and eventually could not alert anybody during an emergency. Another participant recalled how a relative was "*determined to live in her house*" even though she had multiple falls and injuries. This person did not like any carer or a wearable technology because of the fear that someone else would be taking decisions for the individual which would overall affect the individual's independent living. However, there was a genuine concern about how safe one is while living independently, as one participant recalled an incident about his wife (who was also participating in the discussions) saying, "*she collapsed with stroke a few years ago and lay there for like 12 hours before anyone found out*".

Sub-theme 2.2 : Safety or privacy

Family members and often individuals themselves see the need of safety as paramount importance especially when one is living alone. A participant recalled about an old relative who had multiple falls but was "*too proud to call for help*". All the participants agreed that a monitoring system should make the monitored person feel safe. The participants agreed that to feel safe for themselves or their family members, they were ready to strike a balance with safety being a priority as compared to privacy. Commenting on the intrusive nature of the visual data and having cameras for a monitoring technology, one of the participants commented, "*most of the surveillance at the moment, there is somebody sitting in front of the screen looking at it, whereas in this case, it is the computer interpreting it*". This comment about cameras monitoring an individual was further supported by another participant who very aptly claimed that "*One might feel that it may invade your privacy. It is about striking a balance between privacy and safety*". Having a monitoring system in place would not only support independent living but also provide a safe living environment by possibly triggering an alarm to alert the concerned authorities during an emergency. Along with this, a monitoring system should also be able to keep a continuous log of daily living and well-being. "*I think I want to be independent and be on my own. So this would be perfect to have somebody to keep an eye on me*".

3.4.2.3 Theme 3 : Handling the information

Visual data can be intrusive and should be treated with confidentiality. However, when the participants realised that there would be just number or statistics extracted from the videos for automated analysis instead of the videos, the participants were more eager to accept such a system. The participants were not too concerned about how the analysis is done by a computer as long as the software is trust-worthy and well evaluated. Most of them were concerned about where the information would be stored, as in locally in their own homes or on a central server. There were mixed opinions about the location of the storage of data. Some preferred to store all information locally while a few were fine with the information being transmitted to a central server as long as they are confident that they are safe. Some believed that the carers should have a first access of the information while some expressed that they are much more comfortable with the family having the first hand access to the information. They were particularly concerned about whom the alerts would be issued in case of an emergency. All the participants readily agreed that a holistic monitoring system should be connected to some sort of an emergency service which can be alerted at the earliest.

3.4.2.4 Theme 4 : Granularity of information

There was a common agreement among all the participants that an abstracted form of processed visual data is much more acceptable for analysis than raw images or videos. Processed information and health data should be securely stored and revealed only to the person assigned by the monitored individual. There should be different levels of abstraction of the information, according to the understanding and the expertise of the person viewing the information. For example, if the appointed person is the carer or a family member, then having an overall behavioural profile of daily living with high level information should be enough; "*The hierarchy should allow the individual being monitored and a person/kin approved by that individual to have control over all data*". On the other hand, the profile should not contain very specific information like what they have been eating, or specifically what the person has been doing. In some cases, the participants expressed that recording some form of the raw data is valid for post investigation specially in very special cases when the family members want to find out when and how a particular emergency has happened for post-diagnosis.

3.4.2.5 Theme 5 : Where and what should be monitored

The participants realized that for an overall monitoring system, there should be at least a single camera in each of the rooms within the house. While most of them were happy to have a camera in their kitchen and living room, some of them wanted the system to monitor individuals all over the house including the bedroom. They wanted that the system should be able to cover most part of the home where an emergency is likely to occur. Concerns were raised particularly in moving around from one place to another and climbing stairs. There was a detailed discussion regarding if cameras should be placed within bathrooms. While one participant suggested that having a camera even in the toilet is acceptable; another one suggested that "*if you have got it outside and it knows that the person is likely to appear within half an hour and if that person has not appeared, then probably there is a problem*". The participants were more comfortable if the data is pre-processed and only certain abstracted visual information are extracted for analysis, with one participant commenting, "*Since you store only abstract information and not the actual footage, one being monitored in the toilet shouldn't be a problem. Its only our idea of privacy needs to be changed*". The concept of monitoring higher abstracted information as opposed to actually identifying activities found greater acceptance though a few of them were comfortable with the system identifying what activity the monitored person is actually undertaking.

3.4.2.6 Theme 6 : Access Control

All the participants wanted some sort of control over the technology that would be monitoring them. Everyone defended their individual priority of having control over how the information is stored and to whom it is made available to. Continuous monitoring for a healthy living is acceptable as long as it is a "*free choice*". Initially, the participants expressed a desire for a control mechanism which could be used to switch on and off the technology. However, it is fascinating to note that they reached to the conclusion that having a switch is not a very good idea.

It would re-iterate the problems of a wearable technology where "*You switch it off and forget*". Another participant also pointed out that "*It is possible to allow the individuals to turn the system off, but here must be some sort of alarm system to make the concerned people aware that the system had been turned off*".

3.4.2.7 Theme 7 : Different needs of different people

The participants did not want a system that they would need to adapt to. There were strong opinions as to how different people live differently and function differently. Interventions with technology which would affect their usual way of life is not an accepted proposition. They wanted the system where it would be possible that the technology is "*tailored to the individual*" choice and preference of individuals. For a monitoring system, there cannot be a 'one-fits-all' solution. To learn the behavioural patterns of an individual, one of the participants pointed out that, "*Before setting up the technology in somebody's home, there would be period of individual monitoring*" and hence the choices of the particular individual can be learnt by the system then. The system should be personalized to such an extent that it can be updated from time to time according to the preference of the monitored individual.

3.4.3 Secondary Outcomes

The secondary outcomes were not the main pointers of discussion. The secondary outcomes are themes which organically developed as an offshoot to the previous themes. Though there were not much discussions on these themes, they formed an essential attribute towards how older adults perceive and accept a monitoring technology.

3.4.3.1 Theme 8 : Terminology makes a difference

The terms and definitions we assign with such a technology have an effect on how people react to it and accept it. As one of the participants responded to a fellow participant's concern about intrusion by saying "*this is monitoring, it is not surveillance, thats incredible*" clearly reflects how people's perception changes with certain terminologies. Also a lot of participants could relate to a camera-based solution for a safe environment but they strongly felt that having a camera at home relays a sense of intrusion and surveillance. The term 'visual sensor' was a more unobtrusive synonym for them.

3.4.3.2 Theme 9 : Legal drivers for such a technology

There was a particular concern among the participants regarding the legal overheads of such a technology. The concern was mostly about the information that is captured and stored in the computer and who would have access to such information in case the user is not in a position to make a decision about one's own well-being. As one participant pointed out "*in some cases, people are stubborn*", and leaving the technology at their control may have potential risks. The participants overall expressed a desire for video-based monitoring given the ethical, social and legal overheads are critically reviewed and it should be "*implemented with informed consent*".

3.4.3.3 Theme 10 : Economic Cost

Though the economics of such a technology was not discussed in detail, the participants did bring up the topic of the cost of such a technology. They wanted to find out whether such a technology should be self-acquired or should be a part of the overall social welfare provided by the state. While they assumed that such a technology would potentially be a very expensive acquisition, they also realized that with time the cost is bound to come down as pointed out by a participant, "*Its like the price of computers*".

3.4.4 Discussions

The participants had a positive attitude towards monitoring technologies and almost all the participants saw potential scope of using a camera for a monitoring technology. They do not mind the movements being captured by a visual-sensor as long as there is an assurance of safety and what a monitoring technology does. Though there were mixed reactions among the participants about video-based monitoring; they were in unison about the usefulness of such a technology. The prototype examples of how computer vision techniques are being increasingly used in some surveillance and other creative industries applications, their initial hesitations to a camera-based monitoring system gave way to a more accepting view of such a system when they realized that the videos captured by the camera are not recorded but analysed automatically by a computer-based algorithm. Even though visual data is very contextual and extremely sensitive, the participants realize that probably they are always under surveillance owing to the CCTV cameras in shops and public spaces and they are extremely positive about using these evolving technologies to provide them with a better and independent living condition. "*If one has a control over the kind of activities being monitored then there is no problem*".

In the analysis of the focus group discussions, the three key concluding points were:

- everyone values their independence and given a choice would like to continue to live in their own homes

- safety and security is of paramount importance with an option for individual choice of the amount of intrusion

- monitoring technologies should be personalized according to the needs of individual users.

By the end of the discussions, the participants were eager to know more about the possibilities and advantages of a video-based system. As mentioned by Strof *et al.* in [65], a threshold accuracy of 80% is needed in recognizing the activities of daily living (ADLs) to detect for long term conditions among older adults. The participants were ready to push the boundaries of their individual privacy if there is a possibility of independent living in one's own dwelling, with one participants confidently claiming, "*We will get used to it*".

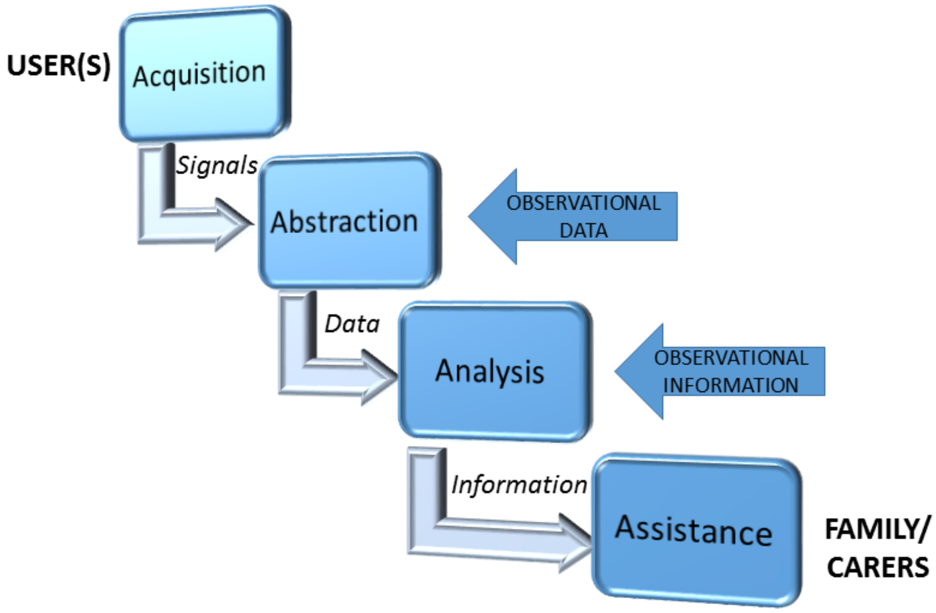


Figure 3.3: 4A's model

3.5 A 4A's model for Video-based Lifestyle Monitoring

The proposed 4A's model as shown in Figure 3.3is a generic model for a visual-sensor based monitoring technology. Each of the four main blocks of the model is detailed in the following 4 subsections. Following the descriptions of the 4 main blocks, a more granular level representation of the model is presented. The 4As model is loosely inspired from the analysis of the focus group discussions and the non-visual sensor based model proposed in [66] and [67].

3.5.1 Acquisition

The choice and the positioning of the acquisition sensors are of paramount importance for a monitoring system. The detection envelope should be such that it covers maximum information from within the room and about the user. The acquisition sensors can be divided into two broad categories, one which monitors the user acquiring details of occupancy, movements, interaction with the environment while the other can be information about the ambient environment.

3.5.1.1 Components

Along with monitoring the activity or activity patterns of an individual, it is also necessary to collect and monitor that ambient parameters for living environment. Ambient parameters like temperature, air quality are essential for a healthy and safe living and should also have a functionality to issue alerts in cases of emergency. For a complete monitoring system, one or more sensors are required in multiple locations. There should be an intelligent switch board which would act like a control hub capable of controlling the activation and deactivation of the sensors based on the location the user is currently occupying. A central hub for controlling the system is required for not only saving energy but also to override manual interventions when needed. As it has been observed in the participants' responses in the focus group, that the users would prefer a control over switching the technology on or off but would also like to have a overriding reminder or alert system which would be activated in case an user switches off the technology and forgets about it thereby making the user vulnerable.

3.5.1.2 Output

Data captured from various types of sensors are mostly different types of signals/electric impulses. These signals need to be sorted and organized into meaningful data which can be analysed into relevant information. For a visual-sensor based system, this would be raw images.

3.5.2 Abstraction

The abstraction layer is used for abstracting out the relevant data from the signals acquired in the acquisition layer. This layer is mostly used to pre-process, sort and aggregate data which can be manually or automatically analysed for more meaningful information about one's living.

3.5.2.1 Input

The input to the abstraction layer is raw signals over a certain period of time. In case of visual sensors, it is images or videos while for the non-visual sensors it would be the sensor firing details.

3.5.2.2 Components

Privacy is of utmost importance for a video-based monitoring system. From the focus group discussions, it was evident that a monitoring system should be for a

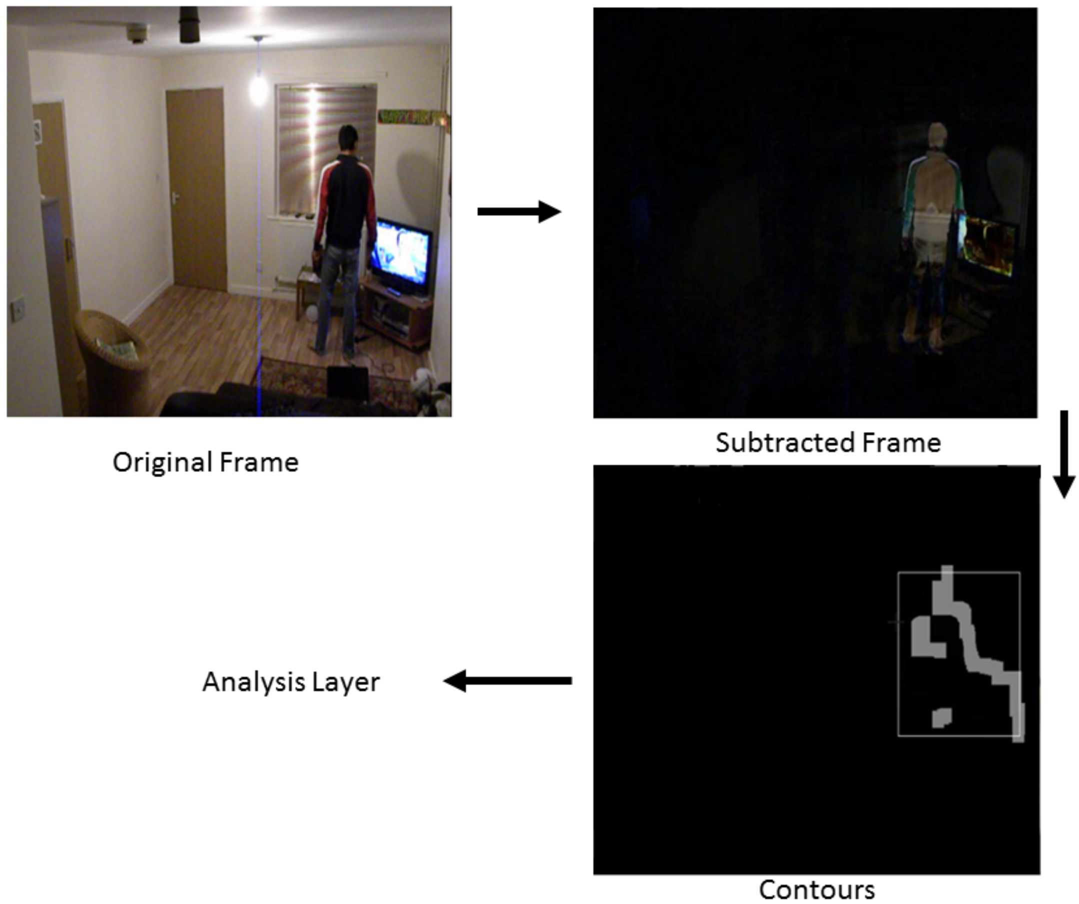


Figure 3.4: Abstraction Layer

better lifestyle and not be like surveillance. Video-data can be anonymized using contours, avatars, 3D models, motion-map *etc.* An example of an anonymized video feed is shown in Figure 3.4. All the participants of the focus groups were ready to accept having a camera at home as long as they had a greater control of the technology and their identity being revealed only to authorized personnels.

3.5.2.3 Output

The data generated in this layer after the primary pre-processing of the raw images can be used as observational data to learn patterns of one's behaviour.

The observational data are mostly for visual analysis experts and health service researchers for co-relating activity patterns and health status and also to develop generic behavioural models.

3.5.3 Analysis

The analysis unit is the technology which processes and analyses the data into information. The amount of analysis needed depends on the architecture of the previous unit. For unprocessed, raw footage this unit would need sufficient computational power to process and analyse the raw video. However, if a portion of the pre-processing is already done in the previous unit, then this unit only needs to analyse the processed data.

3.5.3.1 Input

The input to this layer is the extracted data from the previous layer. This data is then analysed using manual or computerized techniques to give meaningful information. The analysis can yield trend-lines of the pattern of activity undertaken or can be definitive recognition of activities or activity levels.

3.5.3.2 Components

One of the common behavioural patterns which can be monitored using a monitoring technology in an ambient assisted living environment is the mobility pattern. Keeping the participant's responses in mind and the need for having a more abstracted information of daily activities, this unit processes the abstracted data into meaningful information. For this work, a new metric named activity level has been defined (more details in Chapter 4). Activity Levels collected over a period of time is indicative of one's behaviour profile. The profile could be daily, weekly or monthly depending on the choice and needs of the individual user. The profile thus generated can be a part of the observational information the health-service researchers can use for behaviour analysis and identifying trends of one's behaviour.

The meaningful information in this layer can also be used to detect events. An event can be defined as any occurrence which doesn't fall in the normal trend of one's living. It can be an anomalous activity level, or any anomalous recording of the ambient parameters or appliance usages from time to time.

3.5.3.3 Output

The output of this layer is the observational information which can be used by carers to know about the health status and previous records of the user.

3.5.4 Assistance

The two principal stakeholders in a monitoring system are the monitored person (user) and their respective family or carer. The assistance layer has multiple functionality in providing a safe and independent living environment. It can act as a simple reminder service for medication or other activities for the user. Alternatively, it can also host information about the behavioural patterns of an individual derived from the previous layers. This layer can also be used to generate alerts during an emergency (like falls/deviant activity) for which an active assistance is needed. For non-emergency situations, assistance can be provided passively in the form of remote support.

3.5.5 A Granular representation of the 4A's model for Lifestyle Monitoring

The 4A's model is a high level representation of a video-based monitoring system. However, to continuously monitor an individual for a longer period of time, a overall lifestyle monitoring system is needed. "*Lifestyle monitoring systems constitute a sub-set within the wider and more general model of telecare which set out to provide information on a monitored individual's behaviour patterns*" [68]. The telecare systems currently being used are mostly responsive systems which respond to an emergency only after it has occurred and help has been asked for. However with the increase in the elderly population and better living condition, we would be needing the predictive monitoring system which would enable to predict the possible emergency situation by keeping a record and automatically analysing the behavioural patterns of an individual. The goal of a lifestyle monitoring also referred as behavioural monitoring can be broadly defined as monitoring the activities or actions of an individual in their daily living to develop a personalized behavioural profile. Based on the 4As model, a detailed model for lifestyle monitoring is shown in Figure 3.5.

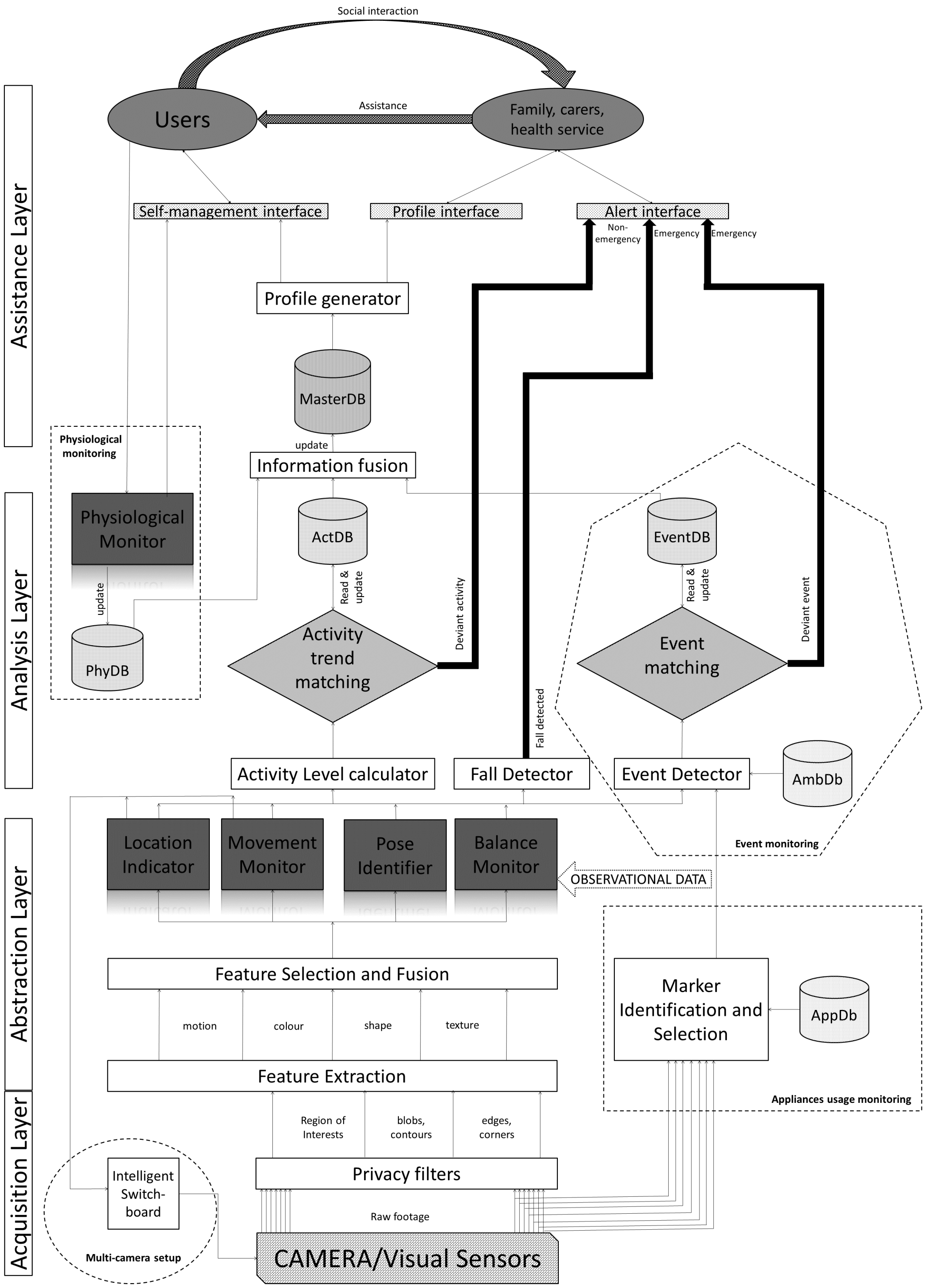


Figure 3.5: Visual sensor based Lifestyle Monitoring System Architecture

3.5.5.1 Main Components

- **Activity Level Calculator**: The activity level calculator estimates the activity levels and maintains a record of the user in the 'ActDB'. In this work, the estimation of activity levels using visual data is presented.

- **Fall Detector**: Older adults are most prone to falls. Hence a fall detector is essential for any monitoring system which has a vision for providing a safe and secure environment.

- **Event Detector**: This works in conjecture with the activity level monitoring to detect for events or complex activities. It fuses visual information about the user along with the various ambient parameters. It also takes information from the appliance usage monitoring service to obtain more relevant context of the event, thereby predicting the event occurred correctly.

3.5.5.2 Additional Components

These are additional components which can be connected to the system based on the requirements of the monitoring technology.

- **Physiological monitoring**: It comprises of a physiological monitor and a database for monitoring and recording of information.

- **Appliance usage monitoring**: This is to monitor the usage of the different appliances within the dwelling. With a prior knowledge about the position and the marker of each appliance, the usage can be effectively monitored and controlled.

- **Multi-modal setup**: For a multi-modal setup within the assisted environment, an intelligent switch is needed which would control the switching on and off of the relevant camera/visual sensor.

3.5.5.3 Databases

There are two types of database in the architecture. One is suffixed as 'DB' and the other as 'Db'. The main difference between these two is the ones suffixed with 'Db' do not contain any information about the users. The type and sensitivity of information within the databases are listed in Table 3.4 and the description is listed below-

- **MasterDB**: This is the master database containing the complete record of the user. This database is updated by fusing the information from the three component databases (PhyDB, ActDB, EventDB). From this database, one can analyse and deduce the behavioural patterns of an individual. The user profile report is created on the basis of this database. Direct access to this database can be given only to authorized personnels.
- **PhyDB**: Contains physiological information of the user. There are several commercially available devices for monitoring blood pressure, glucose count, diabetics and other information. This database is updated through the 'Physiological Monitor' to keep a daily record of the user. The information of this database can be viewed with the 'self-management interface'
- **ActDB**: Contains the activity level information of the user. It also stores information about the different postures of the user. It is updated and

Table 3.4: Database description

Database Name	**Information Type**	**Information Sensitivity**
MasterDB	Behavioural Information	High
PhyDB	Physiological Health data	High
ActDB	Activity Information	Medium
EventDB	Events Information	High
AmbDb	Ambient parameters	Low
AppDb	Appliance	Low

accessed by the trend matching component. Based on the previous records, the technology continuously learns about the various trends of the user thereby adapting the baseline for an alert from time to time.

- **EventDB**: Contains information about the various events within the dwelling. It is updated and accessed by the 'event trend matching' component. It contains not only user information but also other information about the appliance usage and ambient parameters.

- **AmbDb**: Contains information about the ambient parameters like temperature, amount of carbon-dioxide in the air, smoke in the room etc. Most of these information are gathered from various sensors for safety living. Information from this database is continuously fed into the 'Event detector'.

- AppDb: Most of the information in this database is priori knowledge. This database is set-up during system installation with information about the position of furnitures, markers on appliances/doors, unsafe regions of the house. Information from this database is continuously accessed by the 'Marker Identification and Selection' component for appliance usage monitoring.

3.5.5.4 Interfaces

- **Self-management**: This interface is used by the user. It provides graphical representation of the physiological information and the activity levels of the user. This interface can also act as an reminder service for medication or other activities for the user.
- **Profile**: This provides information about the generated user profile from the 'profile generator'. This interface has multi-access controls depending on whether the family, carer of health personnel is viewing the profile information. Since the profile generated is from the MasterDB, it might also contain very specific health-information that can be deciphered or understood only by trained professionals. This interface can also be web-based.
- **Alert**: This is the interface which is invoked when there is a need of assistance to the user. The alert generated can be an emergency situation where active assistance is needed immediately for the user (in case of a fall or deviant event). For non-emergency situations like a deviant activity level, assistance can be provided passively in the form of remote support. Most of the interface is web-based and triggers a relevant alarm depending of the situation.

3.6 Summary

The use of visual sensors for automatic analysis in lifestyle monitoring systems is an evolving research area and has not been investigated widely. While traditional

Table 3.5: Interface Table

Interface Name	Front-end	Back-end
Self-management	Users / patients	Physiological monitor, profile generator
Profile	Family, carers, health personnel	Profile generator
Alert	Family, carers, health personnel	Fall detector, activity trend matching, event trend matching

PIR sensors have often been used for this purpose, visual sensors offer more context to the data obtained as visual data is much richer than usual sensor firings. However, it must also be noted that visual data along with contextually rich, is also extremely sensitive. Moreover, most studies involving visual sensors have mostly been alert systems or used for observational purposes with manual annotating of activities. This is one of the first studies which explores the use of visual sensors for monitoring system for automatically estimating activity level by analysing different visual features. This is done through focus groups, the analysis of which reveals the context of acceptance of visual sensors for monitoring but also have pointed out certain concerns with the granularity of the information that is obtained from the visual data. In the analysis of the focus group discussions, it was concluded that there is a growing acceptance of video-based monitoring subjected to the understanding of the intrusion levels.

Following the results of the focus group discussions a generic model of visual sensor-based monitoring system is presented. The generic 4As model has a high-level representation as well as a granular level representation. While each of the components of the granular level model would require an in-depth study and evaluation, in this thesis, the focus has been on the design and development of the activity level detector in context of the movement monitor. Activity Level as metric for measurement of busyness of an individual from visual data has not been investigated before. While the participants preferred for a higher level of abstracted information about their activities when measured with a visual sensor, there is no reference as to what constitutes an activity level in terms busyness. This thesis first defines activity level through a visual perception test; the results of which are presented in Chapter 4 along with a novel dataset for activity level estimation. Following the definitions of the activity different pixel-based image features are evaluated for activity level estimation; the results of which are presented in Chapter 5 and Chapter 6.

Chapter 4

Visual Sensor-based Activity Level : Definitions and Dataset

In the previous chapter, the context of use and acceptability of a video-based system for assisted living is analysed through focus group discussions. While the context of information obtained from visual data is high, the sensitivity and privacy concerns for a visual information have their own set of challenges. The work in this chapter is to define a metric to measure the busyness of an individual estimated with a visual-sensor and also propose the Sheffield Activities of Daily Living dataset as a novel dataset for activity level estimation.

4.1 Introduction

Within computer vision literature, there is a fair bit of ambiguity over the definitions of an action and an activity. Conceptually an action and an activity are similar to each other but in health service, an action is often referred to as a simple activity like opening a door, turning the oven on or falling whereas activity refers to complex sequences such as cleaning a room or cooking. Action can be defined as a derivative of an activity or conversely it can be said that an activity is made up of several primitive actions.

As shown in Figure 4.1, according to Chaaraoui *et al.* in [1], human behaviour modelling is about a higher level of semantic understanding of an individual's activities or actions. For older adults, these activities usually refer to the Activites of Daily Living (ADL). ADLs (Activities of Daily Living) are categorized into BADLs (Basic Activities of Daily Living) like bathing, dressing *etc.* and IADLs (Instrumental Activities of Daily Living) like housework, shopping *etc.* The BADLs are about fundamental functions of one's living whereas the IADLs tell us more about independent survival [17].

While ADL is the preferred proxy for measuring physical activity among older adults, recognizing activities through a visual sensor may be intrusive. One of the themes of the focus group analysis (Theme 4) emphasized on the need for granularity of information for a monitoring system. Hence, there is a need for a metric which has a higher level of abstraction and not intrusive. In this chapter, activity level is introduced as a metric for measuring the busyness of an individual using visual-sensors. The only study that could be found in the literature using activity levels based on movement of individuals was non-visual sensor-based done

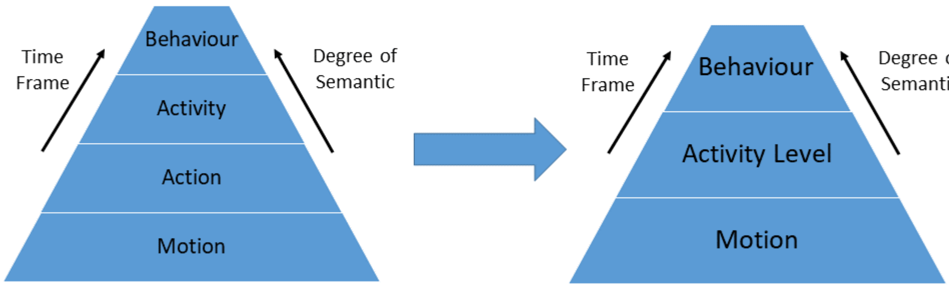


Figure 4.1: Human Behaviour analysis tasks classification - left as proposed in [1]; right as proposed by the author

by Curone *et al.* in [69]. Hence, the definitions of the activity levels are re-defined using a subjective visual perception evaluation. Along with the definitions, a novel dataset is also proposed for activity level estimation. No other publicly available activity recognition dataset have been annotated using the results of a subjective evaluation based on visual perception. The rest of the chapter is arranged as follows, section 4.2 reviews the various available activity recognition datasets proposed as work done within assisted living. Section 4.3 consists of the definitions of activity levels for visual-sensor based estimation. Section 4.4 contains the details of the proposed Sheffield Activities of Daily Living (SADL) Dataset. Finally, a summary of the chapter is presented in section 4.5.

The definitions of activity level for visual-sensor based estimation and the SADL dataset were first proposed in **C1** of 1.1.4.1

4.2 Review of Available Datasets for Activity Recognition within Assisted Living

With the growing interest in computer vision research, there are a number of datasets available to solve different types of problems. Within the action recognition domain, there are a number of datasets that have been proposed over the last decade. Three of the most popular datasets often used for bench-marking of algorithms are the KTH dataset, the Weizmann dataset and the Caviar dataset. The KTH dataset was proposed by Schuldt *et al.* in [70] and comprises of 25 subjects performing 6 different actions (walking, jogging, running, boxing, hand waving and hand clapping) in 4 different scenarios which are mostly outdoor. The Weizmaan dataset first proposed in [71] and later updated in [72] also consisted of 10 actions like run, walk, skip *etc.* performed by 9 actors captured in extremely low resolutions. While both theses datasets had only a single actor in each of the sequences, the Caviar dataset [73] has one or more actors within a sequence carrying out actions like walking, browsing, fighting *etc.* in mostly outdoor locations. Apart from these datasets, some of the activity recognition datasets introduced within the realm of assisted living are discussed later in this subsection.

4.2.1 URADL dataset

The University of Rochester Activities of Daily Living Dataset proposed by Messing *et al.* in [34] is one of the earliest activity recognition dataset with a focus of monitoring activities of daily living, and is widely used as a benchmarking dataset for activity recognition. In the dataset, there are 10 activities simulated by 5 actors within a lab environment. The bottom part of the actor's body is always occluded and all the activities are simulated thrice by the same actor. The position of the actor is fixed for most activities. The most notable flaw of this dataset is that in some of the dataset videos, there is a blurring effect. The videos in this dataset has a resolution of 1280×720 pixels with a frame acquisition rate of 30 frames per second (fps) and is one of the high definition dataset available at the moment.

4.2.2 TUM Kitchen Activity Action dataset

This dataset was proposed by Tenorth *et al.* in [74]. The dataset consists of a single complex activity of preparing table comprising of component activities like taking utensils out of a cupboard, laying the table *etc.* The videos are captured with four fixed overhead cameras with a resolution of 384×288 pixels. Along with the RGB video data, there is also sensor data from the Radio-frequency Identification (RFID) tag readings of the three fixed readers within the environment.

4.2.3 MSR Action dataset

The MSRDailyActivity3D dataset proposed by Wang *et al.* in [36] is created in a controlled lab environment. It has 16 activities by 10 subjects undertaken in a lab environment. All the activities are carried out in a standing as well as sitting posture. The activities are captured using the Kinect camera and hence as an extra channel of depth information along with the RGB channels. Within the dataset, there are a few unrealistic scenarios like using a laptop while standing or the activity 'sit still' in standing posture. Another major flaw in this dataset is even though there is a depth map for each of the activity, the RGB video and the depth map are recorded independently and so they are not synchronized. Also, to compensate for the distance restriction of a Kinect sensor, all the activities are performed very near to the sensor.

4.2.4 Senior Home Monitoring dataset

One of the most extensive and real life dataset is the Senior Home Monitoring dataset proposed by Cheng *et al.* in [35]. The dataset is collected over a period of 4 months within the dwellings of the elderly subjects. There are 6 elderly subjects in the dataset performing 9 different activities like eating, sleeping, drinking *etc.* Though this dataset is extremely extensive and contains a huge amount of data, it is not very good to test algorithms. Also, the subjects are often assisted by carers in carrying out the activities. These sort of dataset can be effectively used for observational purposes.

4.2.5 MPII Cooking Activities dataset

The MPII Cooking Activities dataset proposed by Rohrbach *et al.* in [75] focusses mainly on cooking activities. The database is created within a kitchen where different actions with respect to cooking are simulated. There are about 65 different classes of activities like cut slicing, pouring *etc.* performed by 12 actors in the dataset, captured from a single fixed camera. The resolution of the videos is 1624×1224 with an acquisition rate of 29.4 frames per second (fps). This dataset is not freely available and is licensed to be used only by the researchers within the institute where it has been created.

4.2.6 Discussions

The motivation for the research is to help promote independent living among older adults. Some of the activities mentioned in the above mentioned datasets are not activities of daily living. The rest can be classified as Activities of Daily Living but they are non-representative of the different types of activities. Apart from Senior Home Monitoring Dataset, all the remaining datasets are synthetic datasets simulated under controlled environment with the actor often facing the camera. In all the synthetic datasets, apart from the MPII Cooking Activities Dataset, most of the activities are carried out facing the camera which does not mimic an ideal assisted living scenario. Also, the ground-truth of all the datasets are annotated for activity recognition rather than activity level estimation.

4.3 Activity Level - a metric to measure the busyness of an individual

Hine *et al.* in [76] states that the activities of a person are indicative of his well-being precursors as well as predictive of his well-being outcomes; further elaborating that monitoring of one's daily living should be about visualizing activities at different levels of granularity. According to Gil *et al.* in [18] there are two basic approaches to measure activities of daily living of older adults -

- Monitoring the ADLs by looking at specific activities.
- Measuring the activity or busyness of an individual.

While the first option can be directly referred to as activity recognition, activity level estimation refers to the second option.

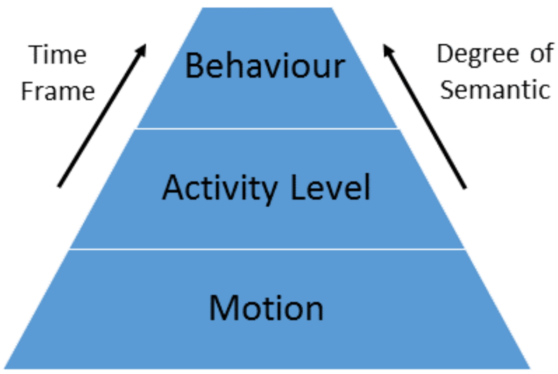


Figure 4.2: Human Behaviour analysis tasks - classification (proposed)

Usual activity recognition has its advantages, but for a visual sensor-based monitoring technology, identifying specific activities undertaken in one's daily living can be intrusive and can lead to privacy concerns. Hence instead of recognizing a specific activity, activity level as a measure of estimating busyness is proposed which would give us enough information about the amount and type of activity an individual has undertaken.

Figure 4.3 shows the work-flow that has been adopted to establish the definitions

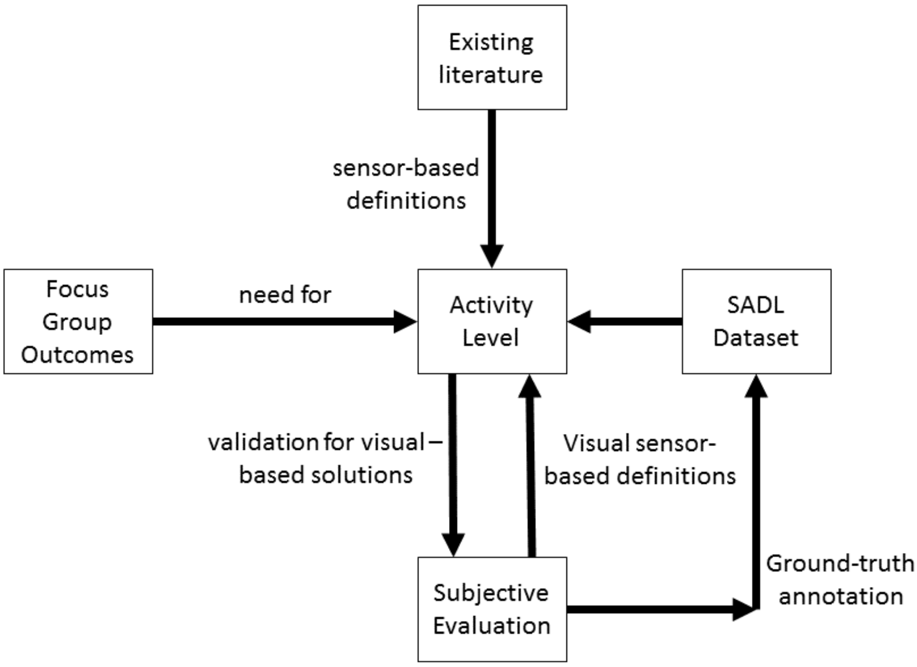


Figure 4.3: Work-flow for this chapter

of activity levels. Activity levels can be defined as groups of ADLs which represent the amount of movement undertaken by an individual in carrying out an activity. It is a measure of the amount of activity one has undertaken. An accurate measure of activity levels over a longer period of time would help model the behavioural pattern of an individual as shown in Figure 4.2. The definitions of the different activity levels are based on the study done by Curone *et al.* in [69], where the authors use a tri-axial accelerometer to determine activities and posture. From the study, it is evident that activity levels can be an effective indicator that would denote the amount of busyness of an individual. These type of definitions are further used by Atallah *et al.* in [?] to categorize activities using a ear-worn wearable accelerometer into 4 classes.

Inspired by the study by Curone *et al.*, the four classes of activity levels were initially defined as follows -

- No - No movement at all
- Low - Minor movement with no change in position within the room
- Medium - Minor movement with a change in position within the room
- High - Major change in position within the room.

Since these 4 categories of activity level were obtained using a wearable sensor, they need to be re-validated for a visual sensor based solution. The validation for these different activity levels were done through a subjective evaluation where videos of activities of daily living were shown and the respondents had to mark the correct activity level.

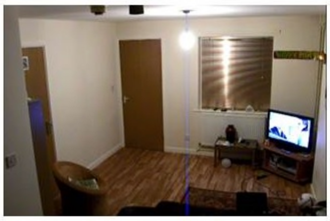
4.3.1 Dataset for Subjective Evaluation

The dataset for the subjective evaluation was created in two different living rooms denoted by 'Living Room A' and 'Living Room B' (Figure 4.4). 'Living Room A' was kept under artificial lighting while 'Living Room B' was kept under natural lighting. There were 3 subjects carrying out 4 activities like walking within the room, watching television, reading a book and ironing. As shown in Table 4.1 these activities were grouped into 4 classes of activity Levels based on the definition stated in subsection 4.3. The activities were captured at 30 fps and each video was 5 seconds long. There were 48 videos created for this dataset. The activities were Standing, Walking, Ironing, Reading and Sitting.

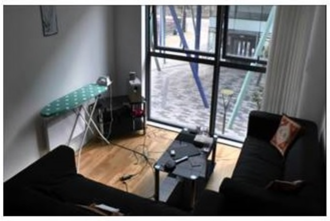
It must also be mentioned that in some scenarios, while standing, the actors were not still and was involved in moving one or more parts of their body.

Table 4.1: Activity Table for initial dataset

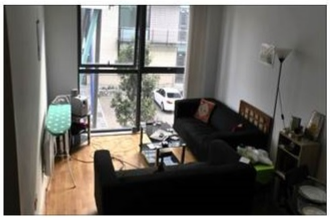
Activity	**Activity Level**
Walking around	High
Ironing and	Medium
Reading or	Low
Watching Television	No



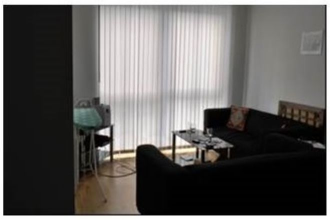
Living Room A (constant viewing angle)



Living Room B (viewing 1)



Living Room B (viewing angle 2)



Living Room B (viewing angle 3)

Figure 4.4: Living Rooms for the initial dataset

4.3.2 Subjective Evaluation Results

The results from the subjective evaluation were used to validate the definitions of the activity levels obtained from existing literature and were tuned for a visual sensor-based solution. The respondents were instructed to label the videos based on what they thought the activity level of the actor would be based on the amount of movement undertaken to carry out the activity based on the definition for each activity level (as defined in subsection 4.3). The responses were collected through a web-based form and are listed in Table 4.2 and Table 4.3. Each of the tables

Table 4.2: Percentage responses for each video in Living Room A

Video No.	**Activity**	**No**	**Low**	**Medium**	**High**	**Final Label**
1	Walking	0	8.33	25	66.67	High
2	Watching	8.33	83.33	8.33	0	Low
3	Watching	16.67	83.33	0	0	Low
4	Walking	0	0	41.67	58.33	High
5	Watching	16.67	75	8.33	0	Low
6	Walking	0	0	33.33	66.67	High
7	Walking	0	0	25	75	High
8	Walking	0	0	33.33	66.67	High
9	Watching	8.33	91.67	0	0	Low
10	Reading	8.33	91.67	0	0	Low
11	Walking	0	8.33	83.33	8.33	Medium
12	Walking	0	16.67	75	8.33	Medium

show the percentage responses for each video which is computed by:

$$Responsepercentage = \frac{Number\ of\ respondents\ marking\ a\ particular\ activity\ level}{Total\ number\ of\ respondents} X100 \tag{4.1}$$

The final label is assigned to a video based on the vote of majority. It can be seen that while there was a majority vote in most of the videos, there are some ambiguous labels. For example for video number 27 and 29, the label is undecided as the responses were equally spilt between High and Medium activity levels. In these two video, the whole subject is not visible and only a part of the upper body is visible. Though by human reasoning, one can establish that the individual is walking, some respondents thought that there is not enough motion present in the scene and hence labelled it as Medium activity level. In video 36, the activity level is marked as low by all the respondents. Even though the activity is sitting, the actor in the video continuously moves his legs or hands which in visual perception is considered to be active. A similar example is seen in video number 15 and 33 where the actor is standing but moving the hands. A graphical representation of the percentages are shown in Figure 4.5.

The results of the subjective evaluation revealed that the boundaries of the different activity levels are fluid and depend on one's understanding of the definitions. The boundaries of the activity levels were not well-defined and differed from user to user based on perception. For example, a couple of respondents did not agree with the term No activity level as according to them even if there is no movement (like reading, watching television etc.), it is still an activity undertaken.

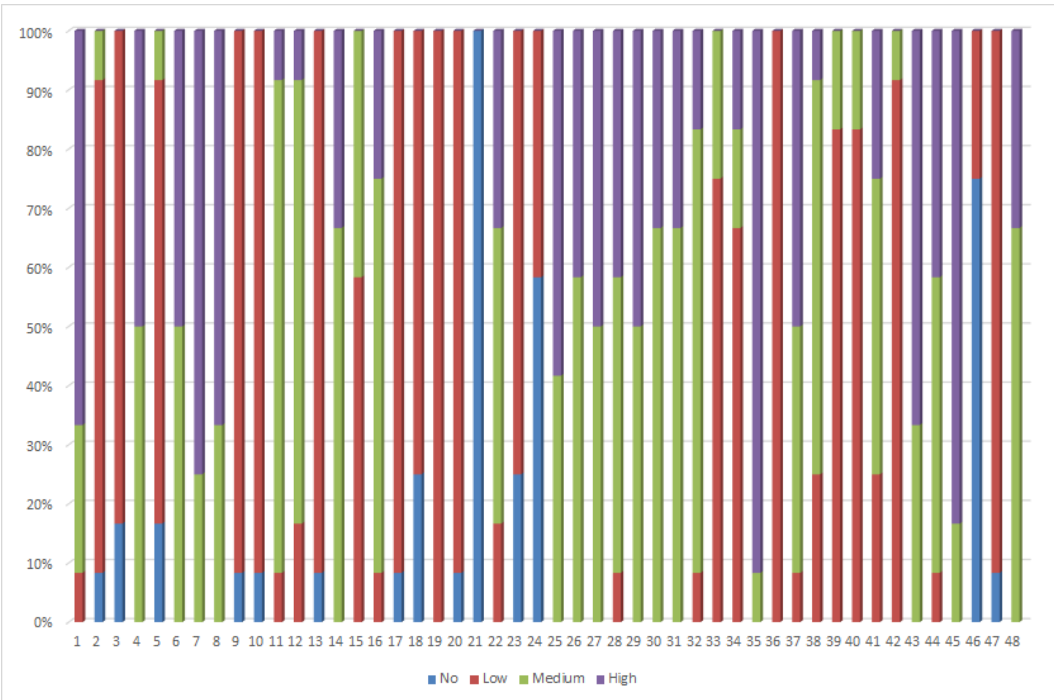


Figure 4.5: Percentage responses for each video

However, taking the vote of majority, the proportion of agreement among the respondents are shown by a pie-chart [Figure 4.6]. The proportion of agreement for most of the videos were over 60%. From the results, the feedback from the respondents, the activity levels for a visual sensor were re-defined and these definitions were used to annotate the ground of the SADL dataset.

Table 4.3: Percentage responses for each video in Living Room B

Video No.	**Activity**	**No**	**Low**	**Medium**	**High**	**Final Label**
13	Reading	8.33	91.67	0	0	Low
14	Walking	0	0	66.67	33.33	Medium
15	Standing	0	58.33	41.67	0	Low
16	Walking	0	8.33	66.67	25	Medium
17	Sitting	8.33	91.67	0	0	Low
18	Sitting	25	75	0	0	Low
19	Reading	0	100	0	0	Low
20	Reading	8.33	91.67	0	0	Low
21	Reading	100	0	0	0	No
22	Walking	0	16.67	50	33.33	Medium
23	Sitting	25	75	0	0	Low
24	Sitting	58.33	41.67	0	0	No
25	Walking	0	0	41.67	58.33	High
26	Walking	0	0	58.33	41.67	Medium
27	Walking	0	0	50	50	Undecided
28	Walking	0	8.33	50	41.67	Medium
29	Walking	0	0	50	50	Undecided
30	Walking	0	0	66.67	33.33	Medium
31	Walking	0	0	66.67	33.33	Medium
32	Standing	0	8.33	75	16.67	Medium
33	Standing	0	75	25	0	Low
34	Ironing	0	66.67	16.67	16.67	Low
35	Walking	0	0	8.33	91.67	High
36	Sitting	0	100	0	0	Low
37	Walking	0	8.33	41.67	50	Medium
38	Walking	0	25	66.67	8.33	Medium
39	Sitting	0	83.33	16.67	0	Low
40	Reading	0	83.33	16.67	0	Low
41	Reading	0	25	50	25	Medium
42	Ironing	0	91.67	8.33	0	Low
43	Walking	0	0	33.33	66.67	High
44	Walking	0	8.33	50	41.67	High
45	Walking	0	0	16.67	83.33	High
46	Sitting	75	25	0	0	No
47	Sitting	8.33	91.67	0	0	Low
48	Walking	0	0	66.67	33.33	Medium

Table 4.4: Summarized Percentage responses for each activity

Living Room	Activity	Activity Level			
		No	Low	Medium	High
A	Walking	0	0	28.6	71.4
	Watching	0	100	0	0
	Reading	0	100	0	0
B	Ironing	0	100	0	0
	Sitting	25	75	0	0
	Standing	33.3	66.7	0	0
	Walking	0	0	58.8	29.4
	Reading	16.7	66.6	16.7	0

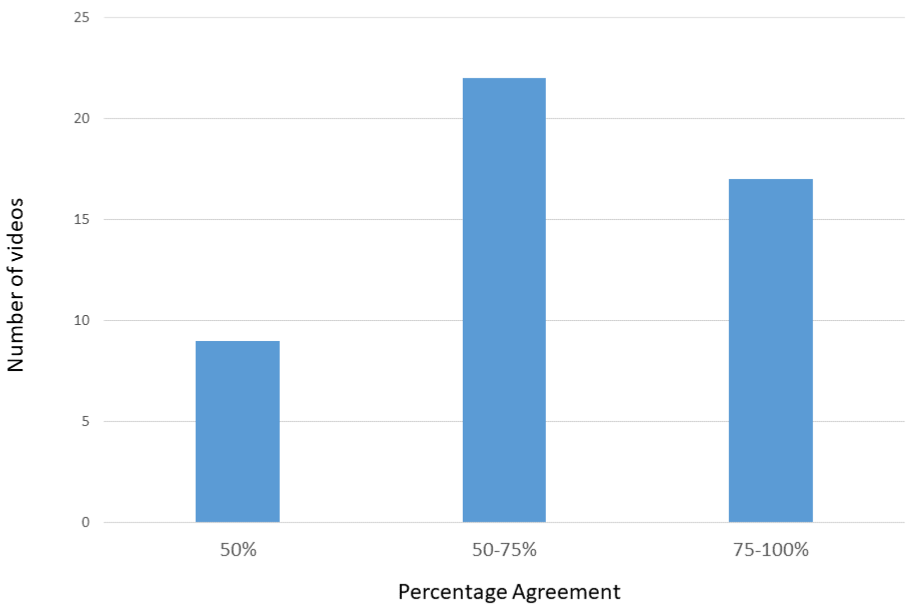


Figure 4.6: The distribution of amount of agreement among the respondents

4.3.3 Activity Levels for Visual-Sensor based solutions

Activity level is a representation of the amount of movement undertaken by an individual in carrying out the activities of daily living. For this work, activity level refers to only physical activity and no other physiological measure. A decrease of Physical Activity Level or physical functionality can reduce the functional dependence and increase the risks of disability with age [13] [14]. Activity Intensity or the energy cost of physical activities have long been manually measured by a physiological measure named metabolic equivalent task (MET). This measure is a ratio of the metabolic rate for a particular activity to the set of reference values defined for that group. The reference values are defined in 3 classes namely light intensity activities, moderately intensity activities and vigorous intensity activities [77]. The recommended activities are mostly physical exercises that an individual should be undertaking daily for healthy living. However for this work, the focus is on activities of daily living which can in-turn fulfil the recommended activity amount and promote independent living.

Based on the previous definitions of activity levels and the results of the subjective evaluation, the 3 classes of activity levels are proposed for this research and are defined as:

- High Activity Level - denotes physically moving from one location to another, *e.g.,* walking,

- Low Activity Level - denotes minor or no change in one's location but involved in some sort of an activity which involves movement or an interaction with the environment, *e.g.,* taking things out of a cupboard or washing plates,

- No Activity Level - denotes extremely low or no movement at all, *e.g.* watching television or sitting idle.

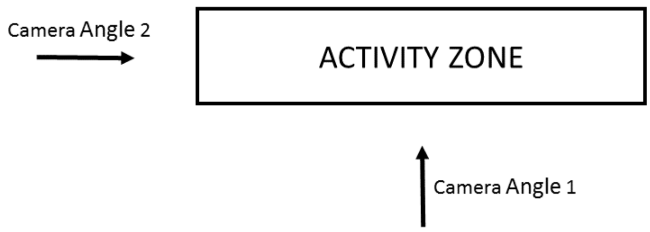


Figure 4.7: Camera Angles for the SADL dataset

4.4 Sheffield Activities of Daily Living [SADL] dataset

While recognizing or monitoring daily activities of living with a visual sensor based system, it can be assumed that most of the activities will be undertaken in an indoor environment. Based on the review of the available datasets for assisted living [section 4.2] and the results of the subjective evaluation, a novel dataset named Sheffield Activities of Daily Living (SADL) dataset is proposed. This dataset is a first of its kind where all the activities are simulated in two different locations within a simulated assisted living environment and captured by two orthogonally positioned cameras as shown in Figure 4.7.

4.4.1 Grouping of Activities into Activity Level

In the focus group discussions, the participants pointed out that the living room and the kitchen are the two locations where an older adult is bound to spend most of the time in a day and carry out most of the activities. Though a holistic monitoring system should be monitoring a person in every location of the house, however for the dataset, only the kitchen and the living room were chosen mostly

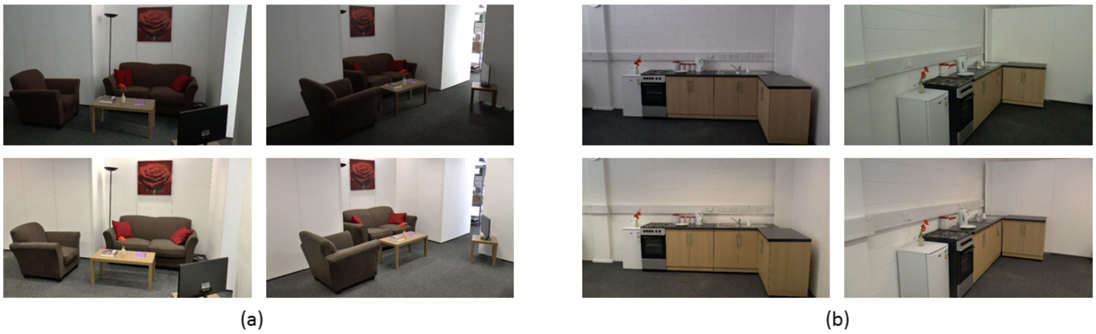


Figure 4.8: Rooms for the SADL dataset

Table 4.5: Activity Table for SADL dataset

List of Activities	**Location**	**Activity Level**
Walking around	Kitchen, Living	HIGH
Using Fridge, Pour water, Using Oven (Plate In and out), Using cupboard, Wash Plate	Kitchen	LOW
Sitting idle, Reading book, Watching Television	Living	NO

because the activities carried out in these rooms are relatively less sensitive to privacy issues than probably using the toilet or sleeping. The location of the dataset videos are shown in Figure 4.8. 9 different basic activities of daily living are identified and are grouped into different activity levels according to Table 4.5. The images for all the activities are presented in Appendix A.

4.4.2 Specifications of the SADL dataset

In the SADL dataset, 9 basic activities of daily living are simulated within Home Lab by 6 subjects. The Home Lab is a simulated indoor environment developed by the Centre for Assistive Technology and Connected Healthcare within The University of Sheffield. The Home Lab mimics the home or care environment and acts as a bridge between the laboratory-based research environment and the homes of people in the community. As shown in Table 4.5, the activities are grouped into 3 classes of Activity Levels based on the amount of movement undertaken while carrying out the activities. The technical specifications of the dataset as follows-

- .avi format,
- video acquisition rate of 50 frames per second (fps),
- resolution of the videos 1920×1080 pixels.

Unlike the other datasets, the activities are not always captured with the subject facing the camera. In the dataset, there are 36 high activity level videos, 60 low activity level videos and 36 no activity level videos for each camera angle making a total of 264 videos. The activities are simulated under two illumination conditions, one being with the indoor lighting switched on and the other being with indoor lighting switched off. The key advantages of the SADL dataset are-

- High definition videos (the highest resolution of all the available datasets),
- two orthogonal perspective of the same activity (as shown in Figure 4.7),
- both the views are synchronized,
- two lighting conditions.

The SADL dataset will be made freely available for download for the research community.

The actors in the dataset have participated voluntarily and the signed consent forms are available in Appendix D.

4.5 Summary

In this chapter, the concept of activity levels as measure for the amount of activity undertaken for behavioural modelling is proposed. Definitions for activity level were initially adapted from available sensor-based definition and re-defined through a subjective evaluation method for visual sensor-based approaches. Physical Activity Level or amount of activity is the visual perception of the amount of movement undertaken while carrying out an activity. Along with the definitions of activity levels, a novel high-definition dataset is also proposed for activity level estimation for further visual sensor-based assisted living research. The ground-truth of the dataset is annotated according to the definitions of activity levels. In the following chapter, an activity level estimation system is proposed and evaluated through different pixel-based motion features.

Chapter 5

Activity Level Estimation using Motion-based Features

Activity Levels are groups of activities of daily living which involve similar movement patterns. Activity levels for home monitoring allow a higher level of abstraction of the information of how much activity an individual has undertaken without identifying the exact activities. The three classes of activity levels were defined in the previous chapter based on the subjective evaluation of visual perception of one's movement undertaken to carry out an activity. In this chapter, an activity level estimation system is introduced and evaluated through various non-intrusive and discriminative motion-based features using two different camera angles. Additionally, a novel motion-moment feature is proposed for activity level estimation.

5.1 Introduction

Activity recognition as is understood in computer vision involves recognizing activities and most of these proposed algorithms are evaluated on datasets where the subject is facing the camera which is not always a realistic representation of an assisted living environment. On the other hand, activity level estimation is about identifying similar groups of similar activities at a higher abstraction to estimate the amount of activity undertaken by an user. The definitions of activity level that is to be estimated by a visual sensor, are defined in Chapter 3 through a visual perception test of the amount of movement undertaken by an individual to carry out an activity.

Activity Level estimation is about detecting similar motion patterns. Human motion is difficult to model as it is articulated unlike rigid bodies. Motion in a video-sequence can be estimated by directly from pixel-based methods or indirectly from feature-based methods. The pixel based methods include optical flow or phase correlation approaches [78], while the feature based methods include tracking of image features like edges, corners or key features like SIFT, SURF *etc.* Though the feature tracking approaches yield good results for the activity recogntion problem, one of the fundamental challenges of feature tracking is the consistency of the tracked key-points [79]. Also, modelling the local descriptor over time is resource intensive and not often discriminatory enough to estimate articulated human motion [31]. Hence, for this thesis pixel-based methods are used to model motion within a scene.

In this chapter, the motion within a video is modelled using the optical flow approach while in Chapter 6 the motion is modelled using phase correlation. Also, since there has been no prior work done on activity level estimation, initially only the motion magnitude and the motion density are used as features and results of these features are used as baseline results to compare the further proposed motion-based features.

The main contributions in this chapter are:

- an activity level estimation system
  - using motion magnitude and motion density as features which forms the baseline results for comparison
  - using motion-HOG features for activity level estimation
- motion-moment - a novel, discriminative hand-crafted motion-based feature for activity level estimation.

The features are evaluated using a single camera setup of the MSR dataset and the SADL dataset along with the dual camera setup {by fusing features from two cameras} of the SADL dataset. Some of the results from this chapter have been published in **C1**, **C2** of 1.1.4.1 and **P3** of 1.1.4.2.

The rest of the chapter is arranged as follows. Section 5.2 details the related work on optical flow methods and motion-based features. Section 5.3 contains the activity level estimation system along with the proposed motion-moment features. Section 5.4 contains the results of the motion-HOG features and the proposed motion-moment features on the SADL dataset followed by the analysis in section 5.5 and summary of the Chapter in section 5.6.

5.2 Background

Motion is an important characteristic in video sequences. It helps to relate the spatial image features over temporal changes. The task of motion analysis is a challenging and fundamental problem of computer vision [80]. Motion analysis has found its application in various applications like robotics, human activity analysis, video compression and video description [81]. Motion in video-data can be defined as a change in a scene, movement of an object in successive frames with a fixed camera angle or movement of both the object along with tilting and panning of the camera. Based on the amount of motion, there can be broadly two categories of motion namely,

- Global Motion - it refers to the motion present in the whole frame. This type of motion can be induced within a frame due to camera movements (pan, tilt, zoom *etc.*) or the movement of an object from one part of the frame to the other part.

- Local Motion - it refers to relatively small changes within a frame where the motion in a scene is in only in a part of the frame.

The movement undertaken by an individual for the low and no activity levels can be categorized as local motion, while that for the high activity level can be categorized as global motion as there is a change across the whole frame.

5.2.1 Optical-flow for Motion Estimation

In the optical flow approach the distribution of moving brightness patterns in an image is estimated. The basic equation of any optical flow problem is given by:

$$w_1 = w_2 + \delta w, \tag{5.1}$$

where w_1 is the pixel in the first frame, w_2 is the same pixel in the successive frame and the displacement is given by δw. Optical flow might be computed densely or sparsely. Dense optical flow refers to movement of image over all pixels in the image while sparse optical flow refers to tracking only interesting points in an image. A popular technique to estimate dense optical flow is the Horn-Schunck method. Under the assumption that "optical flow cannot be computed at a point independently of the neighbouring points" [82], the authors introduce a smoothness constraint along with the brightness constraint to estimate the optical flow. Dense optical flow is usually slow and resource intensive as compared to sparse optical flow. The most popular sparse optical flow method is the Lucas Kanade method. Optical flow techniques have been used to model articulated human motion in human tracking solutions [83] [84] as well as action recognition solutions [85] [86].

5.2.2 Motion-based features

Davis and Bobick had initially proposed the idea of Motion Energy Image(MEI) and Motion History Image(MHI) in [28] to model 'where' and 'how' motion is present in an image. They claimed that the shape of the region affected by motion can suggest about the action and the viewing condition. Given an image $I(x, y, t)$ and the binary image $B(x, y, t)$, the cumulative binary images $E_\tau(x, y, t)$ is be given by:

$$E_\tau(x, y, t) = \bigcup_{i=0}^{\tau-1} B(x, y, t-i). \tag{5.2}$$

These cumulative binary images are called MEIs. On the other hand, to model 'how' a motion is moving in an image they defined a pixel intensity function, H_τ given by:

$$H_\tau(x, y, t) = \begin{cases} \tau \; if \; D(x, y, t) = 1 \\ max\,(0, H_\tau(x, y, t-1) - 1) \; otherwise. \end{cases} \tag{5.3}$$

Some of the key features of MHI are that it can represent motion in a sequential manner and keeps a history of the temporal changes [87]. The MHI was further utilized by Bradski in [88], to have a tMHI template for motion segmentation and pose estimation in which the MHI was updated by:

$$t_\delta(x, y) = \begin{cases} \tau \; if \; current \; silhoutte \; at \; (x, y) \\ 0 \; else \, if \; t_\delta(x, y) < (\tau - \delta \end{cases} \tag{5.4}$$

To get the direction of both the global and local motion in an image, Davis used real time motion template gradients [89]. Davis and Bobick had proposed the idea of Motion Energy Image(MEI) in [28] which is cumulative binary motion images for estimating pose. The authors used the motion features to represent 'where' and 'how' motion is present in an image and also recommended that the shape of the region affected by motion can suggest about the action and the viewing condition.

Apart from these pixel-based methods, tracking of local descriptors is also used to model motion. The HOG descriptor which is one of the most popular object detectors has been used in various applications [90] [91]. However, Dalal (the author who proposed HOG) in [79] shows that tracking the HOG features to model the structure of motion over successive images is not an efficient approach. Instead the author proposed that the HOG features on the motion map obtained from the differential images of the optical flow images is a more robust approach. The differential images are to compensate for camera movements. Activity level estimation is about estimating similar 'type' of motion patterns in a scene. The structure of the motion fields provides important information about the type of motion.

5.3 Proposed Activity Level Estimation System

In this section, an activity level estimation system is proposed. The system works under the hypothesis that pixel-based motion features like motion magnitude and density or features crafted on the motion map can be used to effectively model activity level. The hypothesis is derived from the 3D scatter plot of the motion magnitude and the motion density as shown in Figure 5.1 and Figure 5.2 of all the videos from each of the camera angles of the SADL dataset. The three different colours in the scatter plot show the three different activity levels.

From the figure, it is observed that motion vectors and motion density can be used to discriminate among different activity levels, however for the high activity level and the low activity level, the discriminative boundary is not definite.

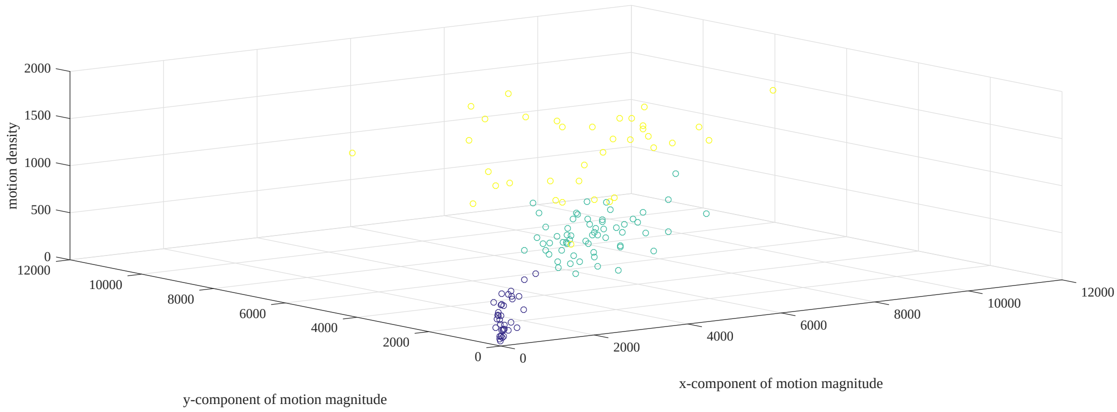


Figure 5.1: Scatter plot of the motion features of Camera angle 1 of the SADL dataset

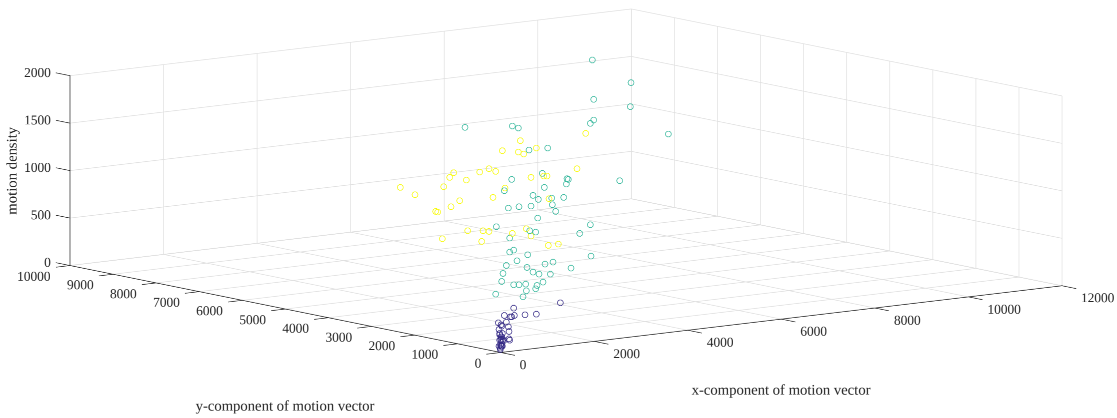


Figure 5.2: Scatter plot of the motion features of Camera angle 2 of the SADL dataset

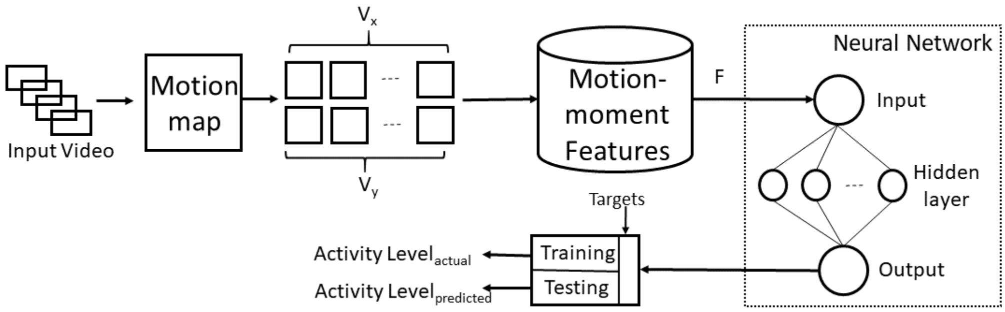


Figure 5.3: Activity Level System setup

To further evaluate the hypothesis, an activity level estimation system is introduced which uses a discriminative classifier to estimate the correct activity level. As shown in Figure 5.3, the system takes in the frames of a video sequence and classifies it to an activity level. The system uses pixel-based motion features as opposed to the traditional feature-based tracking approach for activity recognition. Firstly, a frame-wise motion map is created using the optical flow algorithm, followed by crafting of features in the feature pool and finally classifying them using a neural network. Each of the blocks of the system in elaborated in the following subsections. In the feature pool, the proposed motion features are handcrafted into different variants to recognize the similarities in the patterns. For the initial experiments the existing motion-HOG features are used followed by the proposed motion-moment features for activity level estimation.

5.3.1 Motion Map

In an assisted living environment, it can be assumed that there is no major long range displacement. Hence an optical flow approach is chosen over correspondence based approach as it is more responsive to minor movements and is more sensitive to illumination changes [80]. To compute the motion map, the Lucas

Kanade algorithm for optical flow is used. Lucas Kanade uses the local weighted least squares method to solve the optical flow problem. Assuming that there is constant brightness and there is small motion within each frame, Lucas Kanade proposed an image registration technique using the spatial intensity gradient of the images to find a good match using a type of Newton-Raphson iteration in [92]. The algorithm minimizes the sum of squared error between two successive images and the relative fewer number of matches making it faster as compared the dense techniques. Though proposed almost 3 decades ago, the preformance of the optical flow algorithm in terms of efficiency and accuracy is the best. There have been several implementations of the Lucas Kanade method over the years. In [93], the Lucas Kanade method has been juxtaposed with the Horn-Schunck method to evaluate the local based methods over the dense global methods.

Another implementation of the Lucas Kanade method is the pyramidal implementation proposed by Bouguet in [94]. The pyramidal representation of the image is built in a recursive fashion with I_0 being the original image, I_1 the next level of the pyramid and so on till I_{L-1}. The image at the $L-1$ level is given in the equation:

$$
\begin{aligned}
I^L(x,y) = \\
& \frac{1}{4} I^{L-1}(2x,2y) + \\
& \frac{1}{8}(I^{L-1}(2x-1,2y) + I^{L-1}(2x+1,2y) + \\
& I^{L-1}(2x,2y-1) + I^{L-1}(2x,2y+1)) + \\
& \frac{1}{16}(I^{L-1}(2x-1,2y-1) + I^{L-1}(2x+1,2y+1) + \\
& I^{L-1}(2x-1,2y-1) + I^{L-1}(2x+1,2y+1)).
\end{aligned} \tag{5.5}
$$

The disadvantage of the pyramidal implementation is that it is memory intensive for the generation of the image pyramids. However with a correct selection of the number of levels of the pyramid, this implementation shows good results for real-time calculation of the optical flow algorithm [95] [96]. The efficacy of the real time computation as activity level estimation application is detailed in Appendix C.

For collection of moving images within a time interval of T, the motion map is computed at every time instant t'. Between a pair of frames, $I_1(x,y,t_1)$ and $I_2(x,y,t_2)$ where:

$$t_2 = t_1 + t', \tag{5.6}$$

a motion map ($V(x',y',t')$) is obtained. The motion map is collection of many motion vectors (v^k) given by:

$$V(x'y't') = \{v^k \ : \ k = 1,2,3.....x'y'\}, \tag{5.7}$$

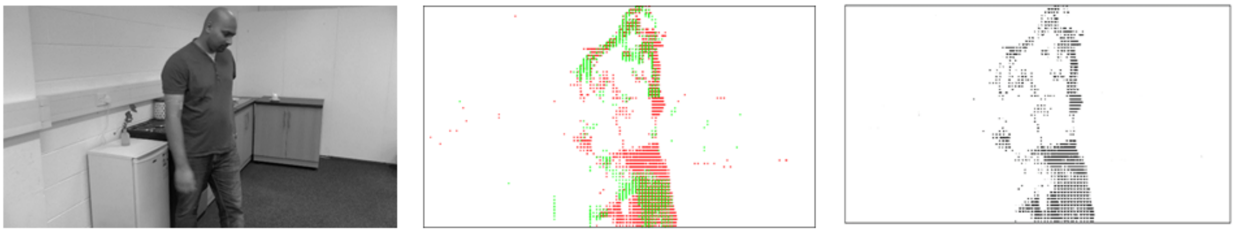


Figure 5.4: Original Frame (left), Motion map (middle), Binary motion map (right)

where each motion vector v_k can be represented as a tuple (v_a, v_b) and v_a and v_b represent the magnitude of the x-components and the y-components of each of the motion vectors respectively. Each motion vector is represented by a magnitude and direction. Since activity levels are an estimate of the amount of motion, hence for this work, only the magnitude is used. A representative motion map is shown in Figure 5.4, where the 'red' lines denote the horizontal motion vectors while the 'green' lines denote the vertical motion vectors. Along with the motion magnitudes, the motion density (v_d) for each $V(x'y't')$ is computed where $v^k > 1$. Two separate motion maps $V_a(x', y', t')$ and $V_b(x', y', t')$ comprising of the horizontal motion vectors and the vertical motion vectors are constructed from $V(x', y', t')$.

5.3.2 Feature pool

In the feature pool, features are hand-crafted on the magnitude of the motion vectors. The features are separately crafted on the horizontal and vertical component. In the pool, the motion-HOG feature and the proposed motion-moment features are crafted.

5.3.2.1 motion-HOG feature

Coupled with a trained SVM, the HOG detector is one of the most widely used person-detector [31]. The standard HOG implementation involves computing the gradients and collecting the weighted votes of the gradients into histograms of an image. The characteristic property of the HOG is that it encapsulates a lot of structural information from a scene. The motion-HOG are computed as proposed by Dalal in [79] to estimate activity levels. In [97], the authors use the motion-HOG features by computing the gradient histograms of the euclidean distance of the x and y components of the motion magnitude and named them as the HOOF features. In this work, the features are computed separately on the two different components.

For each of the motion map, a set of HOG features are extracted. Hence, two set of HOG features H_a and H_b are obtained for $V_a(x', y', t')$ and $V_b(x', y', t')$ respectively. Instead of using the 'winner takes all' strategy as mentioned in [98], the features were further refined by ignoring the regions where there was no motion and quantized by principal component analysis using the single value decomposition algorithm given by h_a and h_b. These sets of l_1 and l_2 principal components of h_a and h_b were considered for the motion-HOG feature vector f_{M_H} given by:

$$f_{M_H} = \{(h_{a_1}, h_{a_1}....h_{a_{l_1}}), (h_{b_1}, h_{b_1}....h_{b_{l_2}})\}. \tag{5.8}$$

5.3.2.2 Proposed motion-moment feature

The motion-moment features fuse the rich invariant properties of Hu moments and the motion features to identify similar ‘type’ of motion which is suggestive of the type of activity undertaken. Image moments have found their application in various object categorization problems but have never been explored on motion features. The motion-moment features are hand crafted features built on low-level motion features based on the concept of image moments. Image moments represent the gross characteristics of an image and are useful for analysis of binary images. The central moments of an image tell us about the shape of an image and are invariant to translation. It must be noted that motion-moments are distinctly different from image moments as used for action categorization in [99], where the moments are computed on the differenced binary image. Motion-moment on the other hand is computed on the motion magnitude values obtained from a motion-estimation algorithm.

Based on the magnitude of the motion vectors, the binary motion vectors b_k are computed as:

$$b_k = \begin{cases} 1: & if\ v_a|v_b > 1 \\ 0: & otherwise. \end{cases} \tag{5.9}$$

Using the binary motion vectors for each of the x-component and the y-component, two separate binary motion maps $B_a(x', y', t')$ and $B_b(x', y', t')$ are constructed. The combined binary motion map $B(x', y', t')$ is the accumulation of all the binary motion vectors and is given by:

$$B(x'y't') = \{b^k \ : \ k = 1, 2, 3.....x'y'\}, \tag{5.10}$$

A representative binary motion map is shown in Figure 5.4. Initially, Hu had proposed 7 Hu moments which are invariant to translation, rotation and skew [100]. In this work the first 4 orders of the Hu moments are computed on each of the binary motion maps to identify the structure of the motion. If each of the binary motion maps $B_a(x', y', t')$ and $B_b(x', y', t')$ is represented as B_R, then the motion-moments thus computed are given by:

$$M_1 = \eta_{20} + \eta_{02}, \tag{5.11}$$

$$M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \tag{5.12}$$

$$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2, \tag{5.13}$$

$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2. \tag{5.14}$$

The scale invariant moment η_{ij} is computed by:

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^{(1+\frac{i+j}{2})}}. \tag{5.15}$$

from the normalized central motion-moment given by:

$$\mu_{ij} = \sum_{x'} \sum_{y'} (x' - \overline{x'})(y' - \overline{y'}) B_R. \tag{5.16}$$

Only the first four order is enough as the remaining three higher orders offer negligible or incoherent information for motion features. Hence the motion-moment features $f_{M_{M_a}}$ and $f_{M_{M_b}}$ for each of binary motion map is $\{M_1, M_2, M_3, M_4\}$ and the final motion moment feature vector (f_{M_M}) is given by:

$$f_{M_M} = \{f_{M_{M_a}}, f_{M_{M_b}}\} \tag{5.17}$$

5.3.3 Classification using Neural Networks

Artificial neural networks are learning algorithms that are inspired by the understanding of how the human brain learns. Neural Networks find their application in practical applications such as speech recognition [101], object recognition, hand written digit recognition, temperature prediction and so on. Neural networks also show promising predictions even for safety critical applications like medical research [102]. Neural networks is a discriminative classifier as opposed to a generative classifier such as Naive Bayes. Discriminative classifiers use posterior probability rather than joint probability as generative classifiers. As mentioned in [103] that the classification problem should be solved based on the specific problem rather than a general problem, there seems to be a wide consensus to the fact that discriminative classifiers perform much better than generative models [104]. Though the training of neural networks is computationally expensive, it requires comparatively less data to predict the interaction between the predictor variables. More recently, neural networks have also found applications in the activity monitoring domain [105] [106].

5.3.3.1 Neural Networks parameters

Owing to three activity level classes, the network has a 3 neuron output layer owing to the 3 different activity level classes and there are 10 neurons in the hidden layer. The network is trained using the Levenberg-Marquardt algorithm with a log-signmoid transfer function. The LM algorithm is one of the most widely used algorithm for optimization with a lower time complexity and higher recognition rates [107].

5.3.3.2 Performance metrics

The classification accuracy for each class of activity level for each of the experiments are presented in the form of a confusion matrix. The True Positive Values in the confusion matrix is normalized over the total number possible positive detections for a particular class of activity level. The performance of the neural networks is presented as a receiver operating curve (ROC). The ROC is a False Positive Ratio(FPR) versus True Positive Ratio(TPR) curve, where for threshold values across the interval [0; 1], the FPR is computed as the number of outputs less than a particular threshold over the total number of false positives and the TPR is computed as the number of outputs greater than the threshold over the total number the true positives. The ROC curve is a representation of the cumulative frequencies to true positives and false positives for a number of thresholds varied between 0 and 1. Hence, for each threshold value, a pair of TPR and FPR is obtained. These classification parameters have been kept constant for all the features.

Table 5.1: Activity Table for the MSR dataset

List of Activities	Activity Level
Walk	High
Drink, eat	Low
Read book, write on paper , sit still	No

5.4 Results

The results of the various different simulations are presented in this section. The simulations are carried out under a single camera setup and a dual camera setup. Under each of the setup, there are 4 different features used as described in subsection 5.4.2. For the single camera setup, along with the SADL dataset, a subset of the MSR dataset is also used. Each of the dataset are divided into mutually exclusive training and testing sets.

5.4.1 Training and Testing subsets of MSR and SADL Datasets

The classification is carried out with a mutually exclusive training and testing set containing approximately 50% of the total number of videos in each set for each activity level. The MSR dataset is annotated as an activity recognition dataset and hence had to be re-labelled according to the definitions of the activity levels. Also, some of the activities within the dataset like using a laptop while standing, play game in standing position *etc.*, are not activities of daily living within an assisted living scenario. Hence, a subset of the dataset is created and the ground truth of the subset dataset is annotated using the definitions as presented in Table 5.1.

For both the MSR and the SADL dataset, approximately $2/3^{rd}$ of the dataset is used as a training set while the rest is used as the testing set. The training and the testing set were kept constant for each of the simulations.

5.4.2 Features Vectors for Simulations

As mentioned in 5.1, there are 4 different combinations of motion-based feature vectors used for the simulations. The feature vector used for each of these simulations are:

1. Feature Vector 1 (F_M^S) = $\{v_a, v_b, v_d\}$
2. Feature Vector 2 ($F_{M_H}^S$) = $\{F_M, f_{M_H}\}$
3. Feature Vector 3 ($F_{M_M}^S$) = $\{F_M, f_{M_M}\}$
4. Feature Vector 4 ($F_{M_C}^S$) = $\{F_M, F_{M_H}, F_{M_M}\}$

All the simulations on the SADL dataset are carried out under two different angles of a single camera setup and a dual camera setup. Under the dual camera setup, the occluded regions and the non-translational motion from one of the camera angle is compensated by the other camera angle. Both the camera angles are equally biased thereby contributing equally to final feature vector. If there are q_1^S features of the feature vector F_1^S from camera angle 1 and q_2^S of the feature vector F_2^S from camera angle 2, then the final feature vector for the dual camera setup F^D is given by:

$$F^D = \left\{ q_1^S q_2^S : q_1^S \in F_1^S ; q_2^S \in F_2^S \right\} . \tag{5.18}$$

Table 5.2: Baseline Confusion Matrix with motion magnitude and motion density on MSR sub-dataset, SADL dataset [Camera 1], SADL dataset [Camera 2]

	High	Low	No
High	**0.40**	0	0
Low	0	**0.10**	0.20
No	0.60	0.90	**0.80**

	High	Low	No
High	**1**	0.20	0
Low	0	**0.80**	0
No	0	0	**1**

	High	Low	No
High	**0.75**	0.35	0
Low	0.25	**0.65**	0
No	0	0	**1**

For each of the simulations, a separate network was trained iteratively using the specifications mentioned in section 5.3.3 until a minimum gradient was reached. The training state of all the simulations of the SADL dataset are presented in Appendix A. The classification results of the test set is given in the following subsections followed by a comparative analysis of the different features.

5.4.3 Results with Single Camera Setup

Activity level is a representation of one's amount of activity. Owing to lack of approaches for activity level estimation, initially the baseline results were obtained with a feature vector comprising of using only motion vectors and classified into different classes using a neural network. Under the single camera setup, the simulations are carried out on the MSR dataset (with a revised annotated ground-truth) and the SADL dataset. The baseline results as presented in Table 5.2 are to test the hypothesis whether motion vectors from optic flow can be effectively used for activity level estimation and also to provide a reference for comparison for further proposed motion-based features.

Following the baseline results, the results with the motion-HOG feature vectors is presented Table 5.3. From the results, it is observed that under a single camera setup, the results are heavily dependent on the positioning of the camera. Under

Table 5.3: Confusion Matrix with motion-HOG features on MSR sub-dataset, SADL dataset [Camera 1], SADL dataset [Camera 2]

	High	Low	No
High	**0**	0	0
Low	0.20	**0.60**	0.50
No	0.80	0.40	**0.50**

	High	Low	No
High	**0.58**	0	0
Low	0.42	**1**	0
No	0	0	**1**

	High	Low	No
High	**0.75**	0.05	0
Low	0.25	**0.95**	0
No	0	0	**1**

Table 5.4: Confusion Matrix of motion-moments features on MSR sub-dataset, SADL dataset [Camera 1], SADL dataset [Camera 2]

	High	Low	No
High	**0.40**	0	0
Low	0.60	**0.50**	0
No	0	0.50	**1**

	High	Low	No
High	**0.83**	0.25	0
Low	0.17	**0.75**	0
No	0	0	**1**

	High	Low	No
High	**0.42**	0	0
Low	0.58	**1**	0
No	0	0	**1**

Table 5.5: Confusion Matrix with all the 3 features combined on MSR sub-dataset, SADL dataset [Camera 1], SADL dataset [Camera 2]

	High	Low	No
High	**0.20**	0	0
Low	0.60	**0.90**	0.85
No	0.20	0.10	**0.15**

	High	Low	No
High	**0.83**	0	0
Low	0.17	**1**	0
No	0	0	**1**

	High	Low	No
High	**0.58**	0	0
Low	0.42	**1**	0
No	0	0	**1**

camera 1, the activity level with the maximum amount of motion is not always classified correctly. However for 'Low' activity levels, the motion-HOG features perform better than the baseline results. The 'Low' activity level is a group of activities which are complex and involves an interaction with an object.

The results show that the motion-moments features, as presented in Table 5.4 are better than the baseline results and are more consistent than the motion-HOG features for the bench-mark MSR dataset as well as the SADL dataset. Also the complexity of computation of the motion moment feature is much less than the motion-HOG feature. However, it must be noted that the performance of the camera 2 angle of the SADL dataset is poor for both the motion-HOG features and the motion-moment feature which suggests that the camera angle plays an important parameter for selection of the feature vector.

Table 5.6: Confusion Matrix for the Dual Camera Setup of Baseline features and motion-HOG features

	High	Low	No
High	**0.83**	0.25	0
Low	0.17	**0.75**	0
No	0	0	**1**

	High	Low	No
High	**0.67**	0	0
Low	0.33	**1**	0.08
No	0	0	**0.92**

Table 5.7: Confusion Matrix for the Dual Camera Setup of motion-moments and all 3 features combined

	High	Low	No
High	**0.92**	0.10	0
Low	0.08	**0.90**	0
No	0	0	1

	High	Low	No
High	**0.83**	0.15	0
Low	0.15	**0.85**	0
No	0.02	0	1

5.4.4 Results with Dual Camera Setup

In this section, the experiments are carried out under a dual camera setup. In the SADL dataset, each activity is captured using two orthogonally positioned synchronized camera angles. This was done to investigate so that if the motion from an activity is occluded from one camera, then the motion from that activity could be captured from the orthogonally positioned second camera. The results for the fused features from both the camera angles is shown in Table 5.6 and Table 5.7. However, under the dual camera setup the performance goes down for all the features apart from the motion-moment feature.

5.5 Comparative Analysis

Motion vectors obtained from optical flow form an important cue to detect similar types of activities or activity level. The ROC curves in Figure 5.5 and Figure 5.6 show the classification performance of the neural networks. The training state

Table 5.8: Overall accuracy of each of the features under different camera setups

	Baseline	motion-HOG	motion-moments	3 features fused
MSR sub-dataset	54.3%	45.7%	**77.1**%	37.1%
SADL [Camera 1]	90.9%	88.6%	84.1%	**95.1**%
SADL [Camera 2]	77.3%	**90.9**%	84.1%	88.6%
SADL [Dual Camera]	84.1%	88.6%	**93.2**%	88.6%

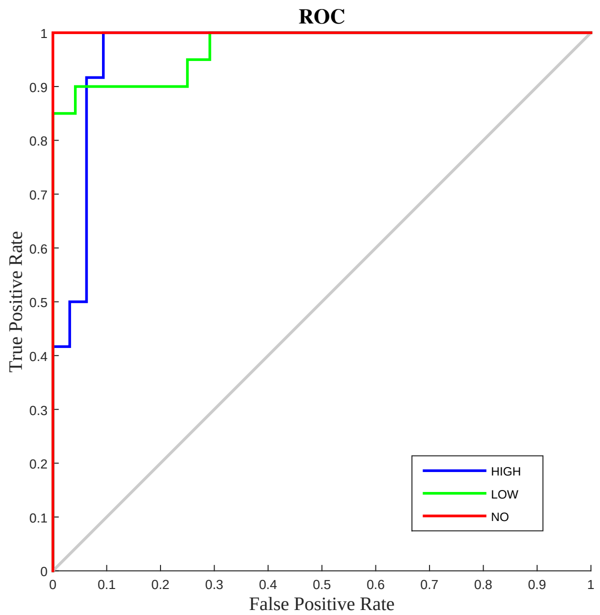


Figure 5.5: ROC plot under single camera setup with Feature Vector (F_M^S) for Camera angle 1

of all the following experiments are presented in Appendix B. The classification accuracy for each class of activity level for each of the experiments is presented in the form of a confusion matrix. The True Positive Values in the confusion matrix is normalized over the total number possible positive detections for a particular class of activity level. The performance of the neural network is presented as a receiver operating curve (ROC).

The overall baseline accuracy for Camera 1 is 90.9% while that of Camera 2 is 77.8% which shows that the positioning of the camera also plays an important role when detecting motion or patterns of motion. To gain more information about

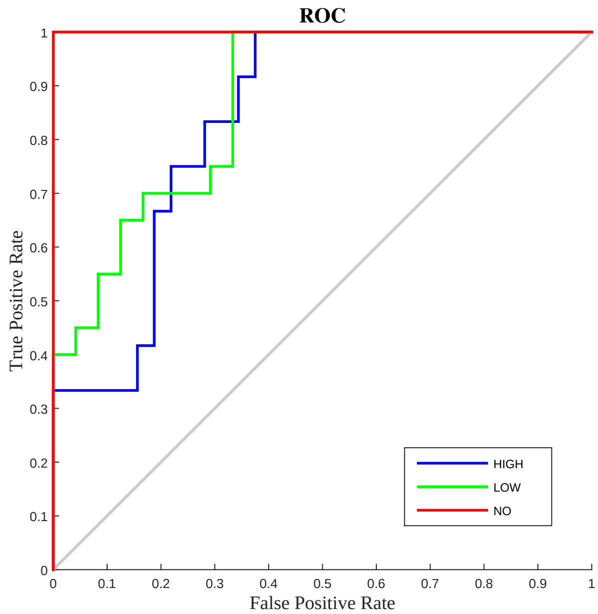


Figure 5.6: ROC plot under single camera setup with Feature Vector (F_M^S) for Camera angle 2

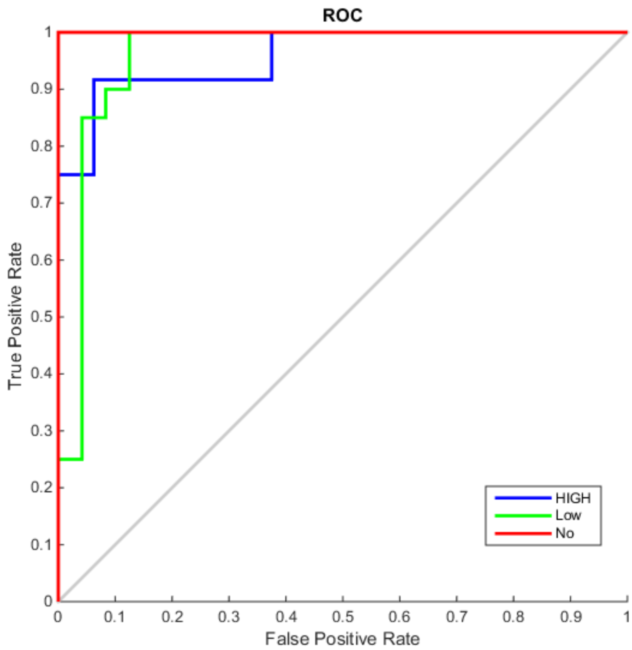


Figure 5.7: ROC plot under single camera setup with Feature Vector ($F_{M_H}^S$) for Camera angle 1

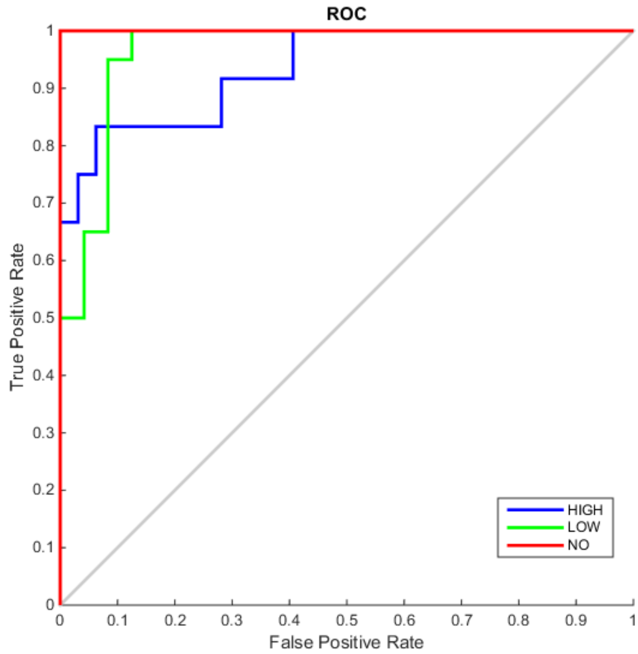


Figure 5.8: ROC plot under single camera setup with Feature Vector ($F^S_{M_H}$) for Camera angle 2

the type of motion, initially the motion-HOG features were used. As shown in Figure 5.7 and Figure 5.8, the performance of the features is comparable with the baseline results for a single camera setup. The motion-HOG features do not perform as good as only the motion magnitude. The number of feature points of the motion-HOG feature vector after reducing its dimension is over 100. Since the parameters of the neural network have not been changed for any of the features to have a comparison among the features, it is possible, the higher dimension of the feature vector causes over-fitting of the model thereby affecting performance. The motion-moment features {Figure 5.9 and Figure 5.10} are crafted from the concept that moments exhibit structural similarities with an image. Motion moments, unlike image moments are not computed based on the pixel values. Instead, motion moments are computed on the motion map comprised of the motion vectors. The motion-moment features were crafted to exploit the use of the invariances that these features offer. With the benchmark

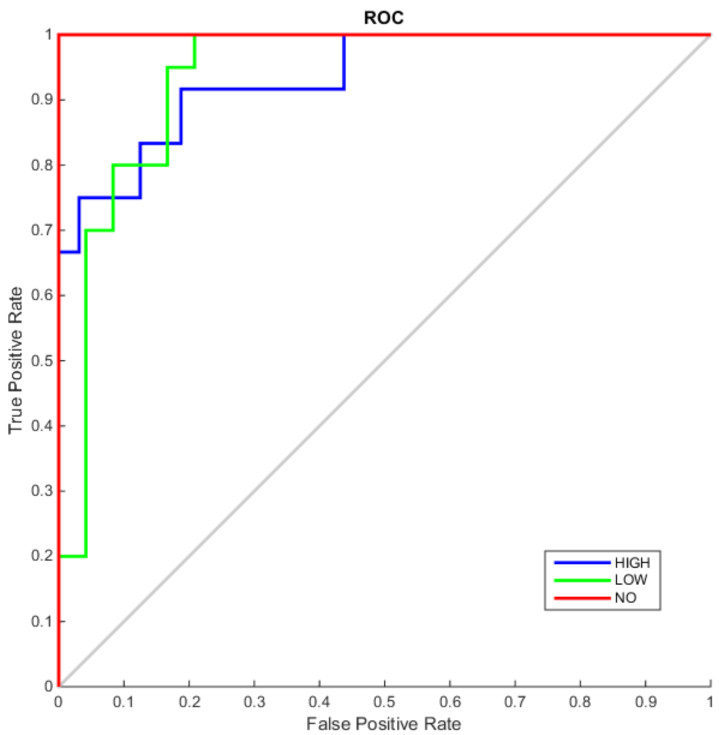


Figure 5.9: ROC plot under single camera setup with Feature Vector ($F^S_{M_M}$) for Camera angle 1

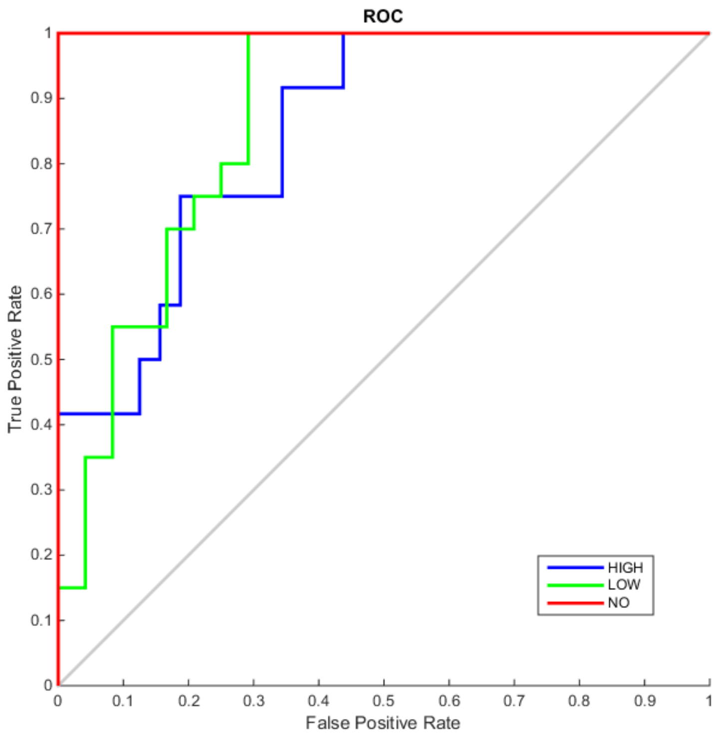


Figure 5.10: ROC plot under single camera setup with Feature Vector ($F^{S}_{M_M}$) for Camera angle 2

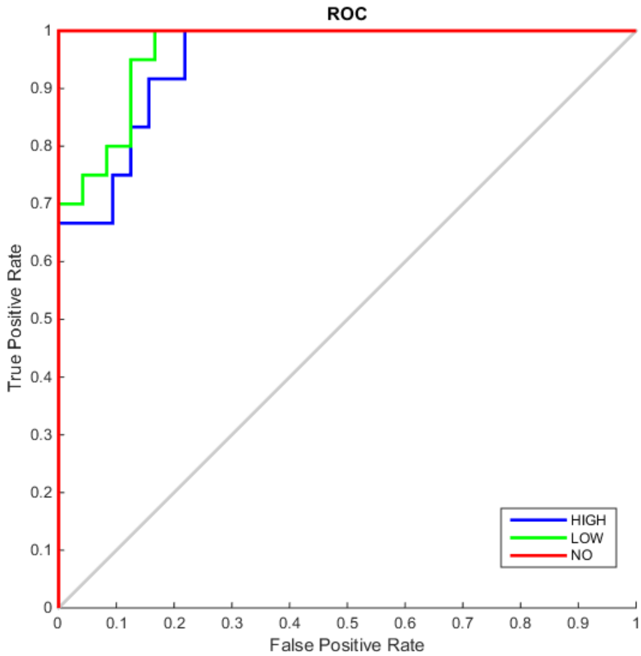


Figure 5.11: ROC plot under dual camera setup with Feature Vector $((F^S_{M_H})$

MSR dataset, the motion-moment feature clearly out-performs the other two features. All the activities in the MSR dataset are facing the camera. However, in the SADL dataset not all activities are facing the camera.

Introducing the second perspective with the motion-HOG features should have improved the classification accuracy even more as it reduces self occlusion and increases the detection envelope within a scene. However, it is observed that there is a decline in the accuracy of the dual camera setup with the motion-HOG features as shown in Figure 5.11 which can be attributed to the over-fitting of the training model due to high dimensionality. However the proposed motion-moment features perform better than all the other features under the dual camera setup. Under this setup, the motion moment not only classifies each activity level more accurately than the other features, but also the overall accuracy of the system is 93%. The proposed feature vector not only exploits the discriminative nature

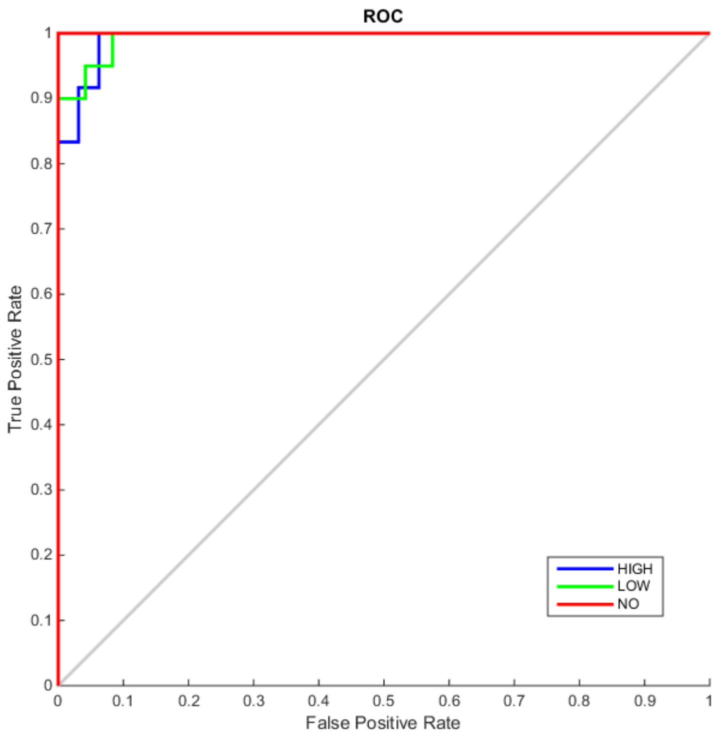


Figure 5.12: ROC plot under dual camera setup with Feature Vector ($F^S_{M_M}$)

of the motion vectors but also effectively indicates the type of the motion more robustly.

5.6 Summary

Motion is an important characteristic in video sequences and it invokes a strong perception. While tracking of key-points and testing them on datasets where the subject is facing the camera is the most popular approach to activity recognition, it is often observed that the consistency of the key-points between frames is not maintained. On the other hand, motion estimation by optical flow gives a good indication of the amount of change in the scene. Activity Levels are a measure of the amount of motion and the type of a particular activity level is also representative of the type of motion. Given the activities in the dataset are not always carried out facing the camera, there is also the problem of occlusion. Introducing the second orthogonal camera angle was to capture information that the first camera angle might miss out. Both the cameras are equally biased and so contribute equally to the final feature vector. From the experiments, it is observed that classifying only the motion vectors and the motion density is not a robust solution for activity level estimation. In an assisted living scenario, the subject is not expected to face the camera always and also, recognizing specific activities can be deemed as an intrusion to privacy. The proposed activity levels are a balance between maintaining one's privacy and yet provide a safe and secure living environment thereby keeping a check on an individual's health condition.

In this chapter, initially the hypothesis that motion vectors can be used effectively to classify activity levels is tested which forms the baseline results for comparison. Using simple motion vectors and motion density as features, the activity levels are classified using neural networks. The experiments were carried out on an available benchmark MSR dataset and also on the introduced SADL dataset. Experiments reveal that the motion values are good indicators of the activity level. The real time efficacy of using such features are relatively low. As shown in Appendix C the activity levels can be recognized with fairly low computation resources and the efficacy of using only the motion vectors for real life deployment of a behaviour modelling system.

Further in the chapter, a new feature named motion-moments is proposed for activity level estimation. Experiments were carried out with a single camera setup using two perspectives and also a dual camera setup by fusing features from both the camera angles. Motion-moment as a feature for activity level estimation outperform both the baseline results and the motion-HOG features under the dual camera setup with an overall accuracy of nearly 93%. In the following chapter, estimation of activity level is explored in the Frequency domain. Modelling motion in the pixel domain is often outperformed by sub-pixel estimation. Since activity level is primarily dependent on the motion information, further experiments in the Frequency domain were carried to compare the results with the pixel domain under the same classification parameters.

Chapter 6

Phase Feature-based Motion Analysis for Activity Level Estimation

As mentioned in the previous chapter, motion plays an important role in mapping spatial variations over temporal changes. Motion estimation in the pixel domain is often an extended application of the image registration problem. In Chapter 4 and 5, motion in the pixel domain is computed using the optical flow technique which was also proposed as an image registration technique and whose application has been extended to video-sequence for motion estimation. The disadvantage of this type of approach is that the translation shift between two images are modelled more accurately rather than the change in the scene. In this chapter, motion is computed in the Frequency domain as compared to the pixel domain to account for the sub-pixel variations.

6.1 Introduction

Motion estimation in the Frequency domain using the properties of the Fourier transform is being used as a popular approach for computing global translational shifts between images. The phase correlation method has been successfully applied to the image registration problem and has also been widely used in object recognition and other motion-based task. With the fast implementation of the Fourier transform, motion estimation in the frequency domain has the obvious advantage of being invariant to shift, rotation and scale changes and also robust to distortions to shape and geometry. Moreover, image registration in the frequency domain shows comparable and in some cases better results than optic flow [108] [109]. Actual motion in the real world occurs in arbitrary. While linear, smooth motion can be correctly modelled in the pixel domain, to get a real representation of the true motion present in a scene, the phase information in the Fourier space is more coherent and robust. Since activity level is a representation of the amount of movement undertaken by an individual, having a robust representation of the true motion within a scene would help in accurate detection of activity levels. The proposed methodology and some of the results have been published in **C3** of 1.1.4.1. The rest of the chapter is arranged as follows. Section 6.2 details the related works of estimating motion and human activity recognition using phase correlation. In Section 6.3, the novel phase correlation methodology in introduced. Section 6.4 contains the results of the experiments using the SADL dataset under the dual camera setup followed by a comparative analysis in section 6.5 and summary of the Chapter in section 6.6.

6.2 Background

Activity level estimation is dependent on the visual perception of the amount of the activity undertaken to carry out an activity. A combined approach to phase and amplitude with a higher weight to phase information leads to better pattern recognition and image perception [110]. Estimation of motion in the pixel domain for real-world applications is more popular due to the emergence of the faster implementation of different algorithms.

6.2.1 Phase correlation for motion estimation

Visual perception of a moving body should be independent of spatial positions and orientations [111]. Fourier descriptors represent the shapes of an object within an image as a distribution of edge information which bears important cognitive information and is independent of illumination changes. The psychophysical aspects of motion perception has been long investigated within neuroscience. Since the phase information offers a topology of the edge information with an image; from a theoretical point of view, due to the Fourier shift theorem the relative shift between two frames corresponds to the phase shift in the frequency domain. Phase information is being widely used within the video-coding community in HEVC encoders [112] [113], visual saliency [114] estimation and video compression [115]. Even within medical imaging, phase properties are effectively used for detection of image artefact propagation, reconstruction and motion reduction in CT-scan images and fMRI images [116] [117].

6.2.2 Phase correlation in Human Activity Recognition

The implementation of the fast Fourier transform has enabled to explore the use of phase properties of the Frequency domain to model motion more effectively. However, phase information is not explored as much within Human Activity Recognition. Weinland in [30] uses the cylindrical coordinates and Fourier magnitudes centred around the bodies to model motion templates and recognize actions. In [118], the authors used a 3-dimension Fourier transform to spot actions in videos. In [119], the authors use the spectral features of the Fourier coeeficients to model action. In [120], poses are estimated using Fourier Transform. Though phase information has been effectively used for estimating the quality of similar patterns [110], it has not been explored for activity level estimation.

6.3 Methodology

Activity Levels are similar motion patterns defined by the amount of activity undertaken to carry out an activity. Phase information provides a qualitative description of cognitive features in a moving image while being robust to substantial level of distortion [121]. In this methodology, the strong invariant properties of the phase information is investigated to model activity levels in a robust manner.

6.3.1 Fourier Transform

Modelling motion in the Frequency domain often refers to estimate the value of a geometric quantity to better the pixel accuracy. Fourier transform is a complex-valued function of frequency which transforms the signal from the time domain into its frequency domain. For a real image $I(x, y)$ where x and y represents the horizontal and vertical coordinates of the pixel intensity I respectively; the Fourier transform, $F(u, v)$ is given by:

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x,y)e^{-2j\pi(ux+vy)}dxdy, \tag{6.1}$$

where u and v are the spatial frequencies. Since $F(u, v)$, is complex hence it can be represented as:

$$F(u,v) = F_R(u,v) + jF_I(u,v), \tag{6.2}$$

where $F_R(u, v)$ is the real part and $F_I(u, v)$ is the imaginary part of the Fourier image. The magnitude spectrum, $|F(u, v)|$ of the Fourier image can be obtained by:

$$|F(u,v)| = [F_R(u,v)^2 + F_I(u,v)^2]^{1/2}, \tag{6.3}$$

while the phase spectrum, $\phi(u, v)$ can be obtained by:

$$\phi(u,v) = \tan^{-1}\frac{F_I(u,v)}{F_R(u,v)}. \tag{6.4}$$

6.3.2 Phase Offset

From the assumption that individual pixel values are less distinctive than the pattern of the whole data, the phase correlation can be used for multi-spectral or illumination varying registration techniques. Phase correlation is often used for modelling the translational offset between two images. Previously, it was often used to solve image registration problems [122]. More recently, phase correlation techniques have been used for modelling motion among moving images [112]. Given two images $I_a(x, y)$ and $I_b(x, y)$, the Fourier transform of these two images can be represented by $F_a(u, v)$ and $F_b(u, v)$ respectively. The normalized correlation cross-spectrum product, $C(u', v')$ between these two images is computed by:

$$C(u', v') = \frac{F_a(u, v) \circ F_b(u, v)^*}{|F_a(u, v) \circ F_b(u, v)^*|}, \tag{6.5}$$

where $\circ$ is the element-wise product and * represents the complex conjugate of the Fourier image. The correlation coefficient which represents the translational offset (k_o, l_o) is calculated by finding the local maxima of the inverse Fourier of $C(u', v')$ as follows:

$$(k_o, l_o) = argmaxC(u', v'). \tag{6.6}$$

Since local level Gabor like Fourier descriptors are preferred over global descriptors [121], instead of taking the whole phase correlation, a novel segmented local approach is proposed which would enable to model the variations in a scene more effectively.

6.3.3 Proposed Localised Correlation Method

Let I_a and I_b are two images where I_b is a shifted version of I_a given by:

$$I_b(x, y) = I_a(x + \delta x, y + \delta y), \tag{6.7}$$

where δx and δy are shift in the x and y directions respectively. After taking the Fourier transform of both the images, due to the shift property the images would be:

$$F_b(u, v) = F_a(u, v)e^{j(u\omega x_c + v\omega y_c)} \tag{6.8}$$

Therefore a phase difference in the frequency domain will occur due to a shift in the phase spatial domain. By correlation theory, we know that the strength of the relationship between these two images can be obtained by computing the correlation between their phase offsets which in turn would denote the amount of change between two images.

As shown in Figure 6.1, the Fourier transform was carried out for each frame. In the proposed method, correlation between two images is defined as the cross-spectrum product of the phase images. Between two given phase spectrum, $\phi_a(u, v)$ and $\phi_b(u, v)$ of images $I_a(x, y)$ and $I_b(x, y)$, the global phase correlation coefficient $\phi_C(u', v')$ is computed by:

$$\phi_C = \Sigma \phi_a(u, v) \circ \phi_b(u, v)^*. \tag{6.9}$$

In the Frequency domain, due to the loss of spatial information, the frequencies are evenly distributed over space and time. However, the spatial information

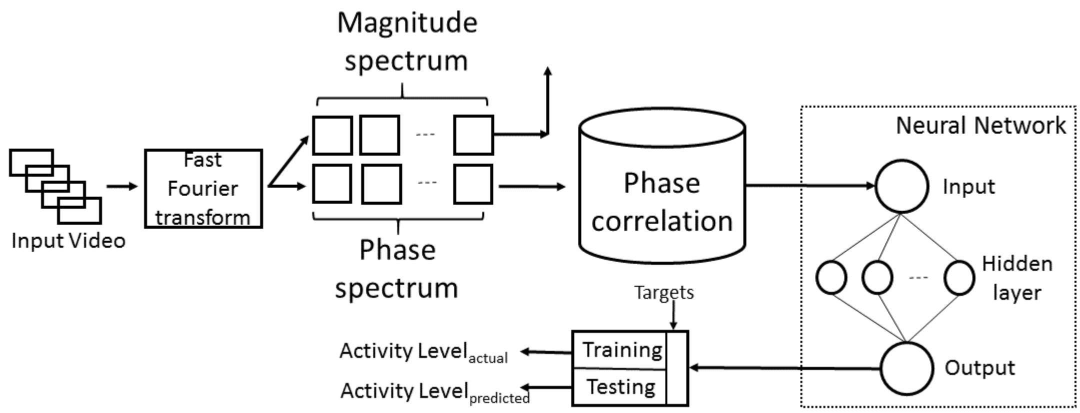


Figure 6.1: Experimental design

loss in the Fourier domain can be compensated by the fact that in activity level estimation, it can be assumed that there would be less or rapid changes to the background of the scene. To have a more localized feature, the correlation coefficients are computed sector-wise. The sector-wise computation of the phase coefficient is inspired by the Gabor orientations. The two most important parameters of the Gabor filters are the kernel size and the orientation. While the kernel size is believed to filter out essential information from a frequency image, the orientations help mimic the human visual cortex. The proposed localized phase feature is computed sector-wise where each sector s is defined as:

$$s = \frac{p^2 sin\theta cos\theta}{2}, \tag{6.10}$$

where θ is orientation angle and p is given by:

$$p = \sqrt{a^2 + b^2}. \tag{6.11}$$

The a and b correspond to the half of the horizontal axis and the imaginary axis of the phase image respectively. Hence, the correlation coefficient is computed

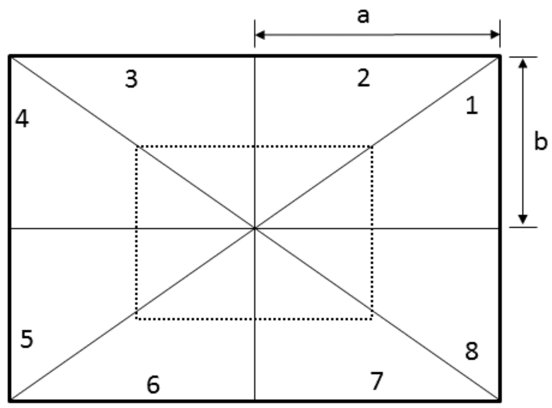


Figure 6.2: Localized sector of a Phase Images $\phi_(u, v)$

as:

$$\phi_{C_s} = \Sigma \phi_{a_s}(u, v) \circ \phi_{b_s}(u, v)^*, \tag{6.12}$$

where s denotes the sector number thereby giving s number of coefficients between a pair of phase images as shown in Figure 6.2. The orientation angle θ is set to 45°.

In a phase image the frequencies are evenly distributed over space and time where, the higher frequencies tend to correspond to mainly noise and are distributed towards the periphery of the horizontal axis, hence the change in the scene is more effectively demonstrated with the correlation coefficients of the lower frequencies. The experiments are carried out using two different sets where the value of a is continually halved before computing the correlation coefficient. In this section, the experimental results are presented. The experimental results are presented in the following section.

Table 6.1: Confusion matrix with localized phase correlation coefficient between $-\pi$ to $+\pi$ and $-\pi/2$ to $+\pi/2$ of the horizontal axis

	High	Low	No
High	**0.75**	0.30	0.17
Low	0.17	**0.60**	0.33
No	0.08	0.10	**0.50**

	High	Low	No
High	**1.00**	0.10	0
Low	0	**0.85**	0.25
No	0.08	0.05	**0.75**

Table 6.2: Confusion matrix with localized phase correlation coefficient between $-\pi/4$ to $+\pi/4$ and $-\pi/8$ to $+\pi/8$ of the horizontal axis

	High	Low	No
High	**1.00**	0.05	0
Low	0	**0.95**	0
No	0	0	**1.00**

	High	Low	No
High	**1.00**	0	0
Low	0	**1.00**	0
No	0	0	**1.00**

Table 6.3: Confusion matrix with full phase correlation and localized phase correlation coefficient between $-\pi/16$ to $+\pi/16$ of the horizontal axis

	High	Low	No
High	**0.67**	0.10	0.17
Low	0.08	**0.45**	0.08
No	0.25	0.45	**0.75**

	High	Low	No
High	**1.00**	0	0
Low	0	**0.70**	0.17
No	0	0.30	**0.83**

6.4 Results

It is known that the frequencies are uniformly spread along the horizontal axis between $-\pi$ to $+\pi$; where the higher frequencies tend to be near the periphery of the phase image, while the lower frequencies tend to be near the origin. Using the SADL dataset under the dual camera setup, in the first half of this set of experiments, all the phase correlation coefficients are used as the feature for classification. In the second half, the phase correlation coefficients between -$\pi/2$ and $+\pi/2$ are computed and used as the feature for classification.

The results for this set of experiments are presented in Table 6.1 and Table 6.2 where it can be observed that the performance increases considerably if we discard the correlation coefficients of the higher frequencies. Higher frequencies

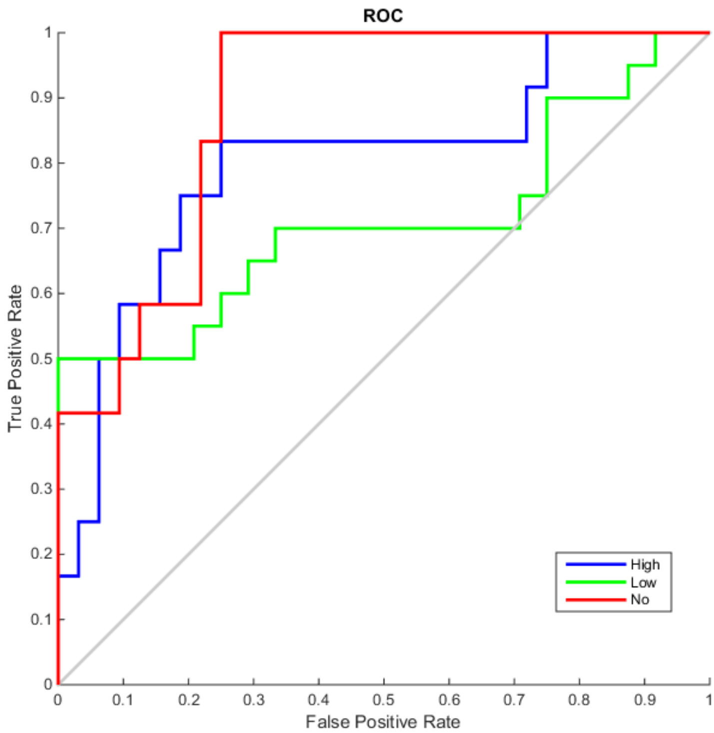


Figure 6.3: ROC curve with phase correlation coefficient between $-\pi$ to $+\pi$ of the real axis

in an phase image often refer to noise and do not contain meaningful and coherent information. The correlation coefficients in this part of the phase image induce errors which in turn affect the classification process. The improvement in the performance of the correlation coefficients of the lower frequency prompted for the second set of experiments to find out the optimal range on the horizontal axis which would be representative of the motion within an image.

As mentioned before, the second set of experiments were carried out to find the optimal range on the horizontal-axis which hold the most meaningful and coherent information in the context of identifying motion. The first half of this set of experiments, the correlation coefficients between $-\pi/4$ to $+\pi/4$ are taken

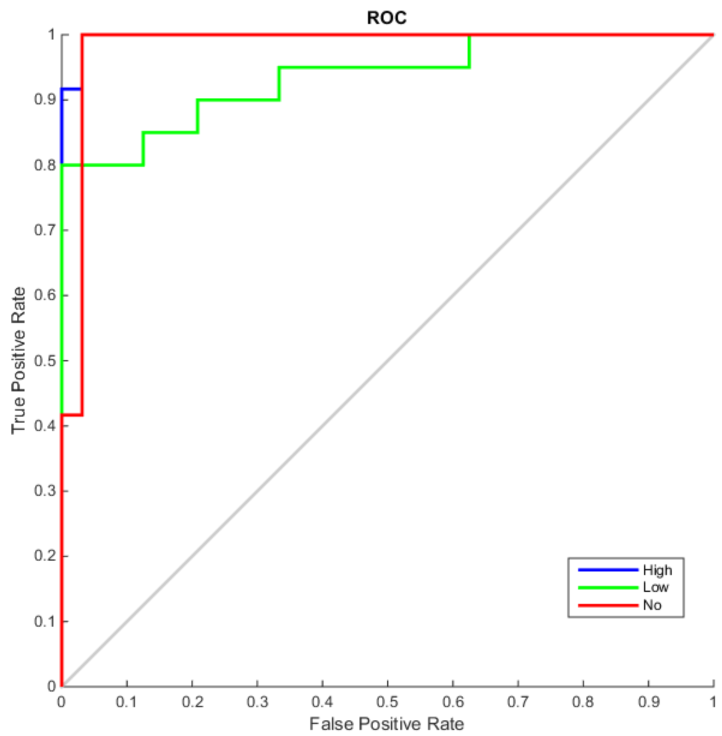


Figure 6.4: ROC curve with phase correlation coefficient between $-\pi/2$ to $+\pi/2$ of the real axis

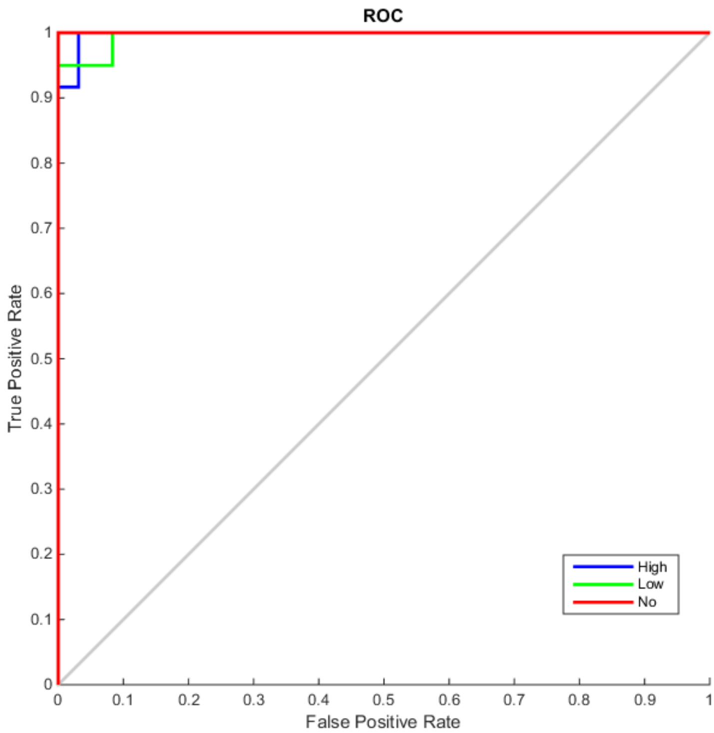


Figure 6.5: ROC curve with phase correlation coefficient between $-\pi/4$ to $+\pi/4$ of the real axis

while in the second set of experiments, the range is halved to $-\pi/8$ to $+\pi/8$.

In this set of experiments, it is observed that performance continues to improve when the range is made smaller. So the correlation coefficients of the lower frequencies hold more coherent information about the type of motion than the higher frequencies. It must also be noted that the higher frequencies do not contribute much in motion estimation. In fact the correlation coefficients of the higher frequencies tend to induce errors in the motion estimation process.

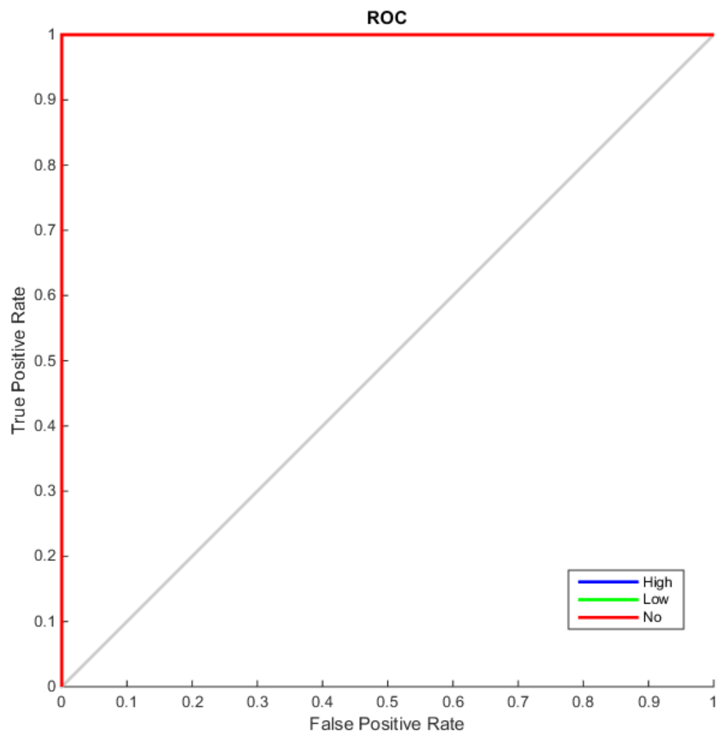


Figure 6.6: ROC curve with phase correlation coefficient between $-\pi/8$ to $+\pi/8$ of the real axis

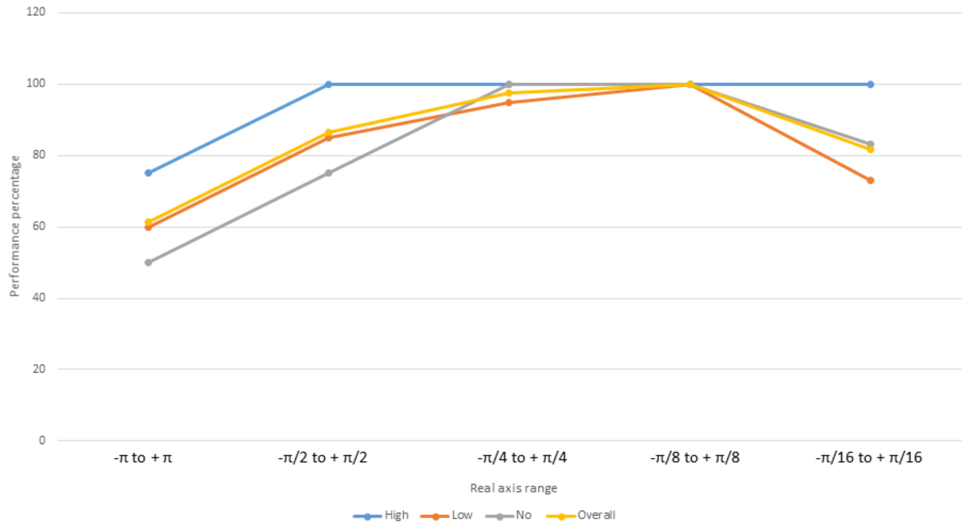
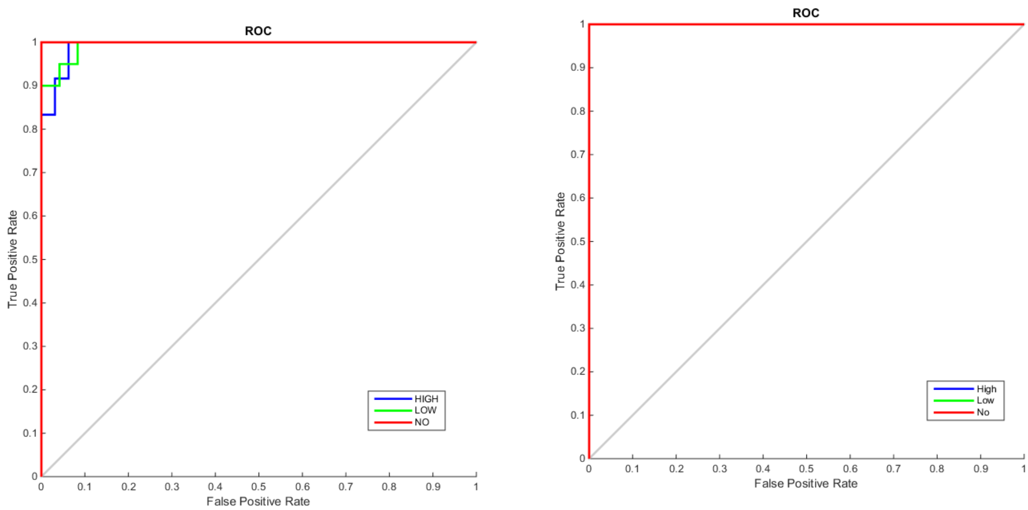


Figure 6.7: Performance comparison of different ranges of the horizontal-axis of phase image

6.5 Comparative Analysis

In the proposed method, the properties of the Fourier shift theorem is explored to model motion. Firstly the proposed method is compared to the traditional global phase shift method. The results of the phase correlation are presented in Table 6.3. From the results, it is observed that the proposed localized correlation features provide more insight about the different patterns of motion present in the video.

The localized features are computed with varying kernel sizes to obtain the optimal range of frequency coefficients needed to model motion. As shown in Figure 6.7, the range of $-\pi/8$ to $+\pi/8$ of the horizontal axis is enough to model similar motion patterns for estimating activity levels. It is observed that even for fewer frequency components, the High activity level still gets classified accurately, however for complex activities, the performance deteriorates.



(a) Feature Vector with motion-moments feature

(b) Feature Vector with phase correlation coeeficient between $-\pi/8$ to $+\pi/8$ of the horizontal axis of 2D phase

Figure 6.8: ROC plot of the best feature in the pixel domain and sub-pixel domain

Comparing the features in the pixel domain and the frequency domain, it is observed that the features in the frequency domain offer more coherent information of amount of motion present in a scene. In the pixel domain, the motion-moments are the most discriminative feature for activity level detection while in the sub-pixel domain, the phase correlation values between $-\pi/8$ to $+\pi/8$ of the horizontal axis are enough for modelling motion for activity level estimation as shown in Figure 6.8.

As shown in Table 6.4, the No activity level where motion is less or negligible has complete accuracy for both the domain. However for the activities with complex motion, the performance in the sub-pixel domain is better than the pixel domain. In the pixel domain, only translational motion is accounted as any other form of change (specially non-translational and rotational) in a scene. On the other hand in the frequency domain, the integer approximation in the pixel domain is eliminated, therefore giving a more accurate representation of the change in a

Table 6.4: Comparison Table between pixel domain and sub-pixel domain

	pixel-domain (motion-moments)	sub-pixel domain (-π/8 to +π/8)
High	0.92	1.00
Low	0.90	1.00
No	1.00	1.00

Table 6.5: Overall performance of all the features for activity level detection

	Performance (in percentage)
pixel-domain (motion-vectors)	84.1%
pixel-domain (motion-HOG)	88.6%
pixel-domain (motion-moments)	**93.2%**
sub-pixel domain (-π to +π)	61.4%
sub-pixel domain (-π/2 to +π/2)	86.4%
sub-pixel domain (-π/4 to +π/4)	97.7%
sub-pixel domain (-π/8 to +π/8)	**100%**

scene.

6.6 Summary

The overall performance of all the features used both in the pixel domain and the Frequency domain is shown in Table 6.5. Under a dual camera setup with both the camera equally biased, the proposed motion-moments features in the pixel domain and the optimal range of frequencies in the frequency domain perform the best.

In this chapter, activity level estimation is carried out using a novel localized phase correlation method in the Frequency domain. Using the phase properties of the Fourier transform to estimate motion, experiments are carried out to estimate activity level using the same classification parameters as the previous chapter. The phase correlation coefficients act as a more discriminative feature for motion estimation for activity level estimation than the proposed features in pixel domain. Experiments show that the phase correlation coefficients of only the lower frequencies ($-\pi/8$ to $+\pi/8$ of the horizontal axis of 2D phase) are enough for modelling motion for activity level estimation.

Chapter 7

Conclusions

This thesis presents the research of a project which is a part of a multi-disciplinary network within The University of Sheffield whose aim was to explore and investigate novel technological solutions/interventions for older adults. The specific research focus of this project was to investigate the use of visual sensors and propose an effective metric to measure the amount of activity undertaken for lifestyle monitoring systems. The rest of the chapter is arranged as follows. Section 1 details the conclusions to this thesis detailing a brief summary of each of the chapters and section 2 details the possible future directions in which this research can be taken forward.

7.1 Conclusions to the Thesis

The overall motivation for this work is the rising population of older adults and the need to have novel technological solutions/interventions to promote independence among the ageing population. There are several strands to how technology is being used as enabler to better and healthier living for vulnerable individuals. Here, the focus has been on lifestyle/behavioural systems which can be used to monitor the amount of activities undertaken by an individual. Such systems have mostly employed non-visual sensors of activity understanding. The use of visual sensors for lifestyle monitoring has been mostly restricted to observational purposes. The use of the visual data for automatically analysing and inferring information about a scene has not been widely investigated or reported. This is one of the first studies which looks into the use and efficacy of estimating activity levels using visual sensors in daily monitoring of activities among older adults. Visual sensors offer much more contextual information about an environment and its occupants with a rather simpler infrastructure. However, it must also be noted that visual data is extremely intrusive and can lead to privacy concerns. As mentioned in the earlier chapters, this thesis investigates the acceptability of visual sensors for monitoring systems, establishes the metric [activity level] that needs to be estimated and then proposes features for effectively estimating activity levels.

In Chapter 3, the findings of the focus groups are presented. Though vision-based activity recognition or monitoring is a relatively new research domain, it was found out that there is a growing acceptance of a concept of such a technology as long as it provides a safe and independent environment. Some participants did express their concern about privacy, but an overall consensus was that if the ethical, legal and practical issues of a vision-based technology are carefully examined, then using cameras would be an acceptable solution. Though the discussions of the focus groups ranged from the usability of technology in modern lives to the importance of the monitoring technologies, the main focus of the discussion was about how comfortable and safe would they feel about having a camera at home for themselves or a family member. Following the analysis of the focus group, a model for using visual sensors for monitoring is presented. The model has a high level representation and a more granular level representation.

In Chapter 4, the concept of activity levels as opposed to activities is defined using literature and subjective evaluation. One of the major challenges of this study was finding a bench-mark dataset to evaluate algorithms. Publicly available activity recognition datasets for assisted living often defeat the purpose of activity recognition within assisted living. For this work, a new bench-marking dataset was created. The dataset was annotated based on the previously presented definitions of activity levels.

In Chapter 5, the methodology to estimate activity levels are presented. Since this is the first work to estimate activity levels, initially simple features like motion magnitude are used to estimate activity levels. These results were used as baseline results for performance comparison to the proposed features under a single camera set-up as well as dual camera set-up. Initially, the motion-HOG features are used to estimate activity levels, which are marginally better than the results of the motion-vectors. However, the proposed motion-moments feature outperforms the baseline results and the motion-HOG features by a large margin under the dual camera setup.

In Chapter 6, the quest for an optimum feature for activity level estimation is continued. In the pixel domain, computation of motion features is expensive. Also the optic flow only accounts for translational motion. For other types of motion, or to find out the true change in a scene, the Frequency domain offers more information than the pixel domain as it models sub-pixel motion as well. Using the phase properties of the Fourier domain, motion is modelled for activity level estimation. A localized phase correlation method is proposed to map the activity levels and the results are compared with that in the pixel domain.

This thesis emphasizes on finding the best feature for activity level detection using a visual sensor. The experiments are carried out on a novel dataset. While activity level is not the only indicator of one's lifestyle, it definitely forms a non-intrusive metric to determine the amount of activity. Visual data offers more information than just motion which can be explored in the future for a holistic lifestyle monitoring system.

7.2 Future Work

This thesis is a result of an interdisciplinary research study, it offers notable future directions which can be extended to various avenues of research within assisted living domains in health service research and computer vision. Activity Level is just one of the indicators that can be used to contribute to the measurement of the overall behaviour or lifestyle of an individual. While each of the components of the presented granular monitoring system needs a detailed study before designing, here are some of the broad directions that this research can be taken forward in-

- The 4As model:
  - An acceptability study using a mixed methods approach on visual sensor based monitoring systems with focus groups, interviews, questionnaires and surveys.
  - Conduct in home trails to evaluate and address the practical challenges of implementing the 4As model.
- Motion modelling for Activity Level:
  - Propose novel motion tracking algorithms invariant to usual computer vision problems like occlusion, shadow detection etc.
  - Model a dynamic background scene with visual saliency based approach to track a single individual within a multi-people environment. The visual saliency model can also be extended to have an ego-centric vision of one's daily activities.
  - Use an extra channel of information (RGB-D) for approximating the exact location of an individual within a room. With the extra channel of data, more contextual information can be added to model and identify the behavioural patterns of an individual.
  - Evaluate the real time efficacy of estimating activity levels for behaviour modelling on modern digital processing boards.
- Extend the SADL dataset to include more activities of daily living and possibly more classes of activity levels. Though the SADL dataset has been subjectively evaluated and annotated, there is scope to identify more complex activities and classify them according to different activity level based on subjective and technical evaluations.

- Explore the possibility of an adaptive biasing of a multi-camera setup to improve efficiency and computation.

- Using visual analytic to represent and model real-life activity level data. This data can be further utilized and fused with other data for behaviour modelling and anomaly detection. Based on the behavioural pattern, a personalized profile for each user can be established to predict one's failing health condition.

References

[1] A. A. Chaaraoui, P. Climent-Pérez, and F. Flórez-Revuelta, “A review on vision techniques applied to human behaviour analysis for ambient-assisted living,” *Expert Systems with Applications*, vol. 39, no. 12, pp. 10 873–10 888, 2012. xviii, 21, 68, 69

[2] United Nations. Department of Economic and Social Affairs. Population Division, *World Population Ageing, 1950-2050*, ser. Population Studies. UN, 2002, no. 207. 1

[3] F. Cardinaux, D. Bhowmik, C. Abhayaratne, and M. S. Hawley, “Video based technology for ambient assisted living: A review of the literature,” *Journal of Ambient Intelligent Smart Environments*, vol. 3, no. 3, pp. 253–269, 2011. 2, 20, 21, 22, 24

[4] T. Rutherford, *Population Ageing: statistics.* Commons Library Standard notes SN03228UN, UK Parliament, 2012. 14

[5] *2010-Based Statistical Bulletin National Population Projections.* Office for National Statistics, UK, 2011. 15

[6] *Efficiencies in Telecare.* Department of Health Publications, 2010. 15

[7] B. Celler, W. Earnshaw, E. Ilsar, L. Betbeder-Matibet, M. Harris, R. Clark, T. Hesketh, and N. Lovell, "Remote monitoring of health status of the elderly at home. a multidisciplinary project on aging at the university of new south wales," *International Journal of Bio-medical Computing*, vol. 40, no. 2, pp. 147–155, 1995. 17

[8] S. Brownsell, D. Bradley, S. Blackburn, F. Cardinaux, and M. S. Hawley, "A systematic review of lifestyle monitoring technologies," *Journal of Telemedicine and Telecare*, vol. 17, no. 4, pp. 185–189, 2011. 17, 18

[9] S. Hargreaves, "A mixed methods exploration of the relationship between activities within the home and health in older people with heart failure: implications for lifestyle monitoring," Ph.D. dissertation, University of Sheffield, 2014. 17, 25

[10] W. H. Organization *et al.*, "Global recommendations on physical activity for health," 2010. 17

[11] P. T. Katzmarzyk, "Physical activity, sedentary behavior, and health: Paradigm paralysis or paradigm shift?" *Diabetes*, vol. 59, no. 11, pp. 2717–2725, 2010. 18

[12] H. Gokalp and M. Clarke, "Monitoring activities of daily living of the elderly and the potential for its use in telecare and telehealth: A review," *Telemedicine and e-Health*, vol. 19, no. 12, pp. 910–923, 2013. 18

[13] M. I. Frisard, J. M. Fabre, R. D. Russell, C. M. King, J. P. DeLany, R. H. Wood, and E. Ravussin, "Physical activity level and physical functionality in nonagenarians compared to individuals aged 60–74 years," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 62, no. 7, pp. 783–788, 2007. 18, 82

[14] W. J. Chodzko-Zajko, D. N. Proctor, M. A. F. Singh, C. T. Minson, C. R. Nigg, G. J. Salem, and J. S. Skinner, "Exercise and physical activity for older adults," *Medicine & science in sports & exercise*, vol. 41, no. 7, pp. 1510–1530, 2009. 18, 82

[15] E. Biddiss, S. Brownsell, and M. S. Hawley, "Predicting need for intervention in individuals with congestive heart failure using a home-based telecare system," *Journal of Telemedicine and Telecare*, vol. 15, no. 5, pp. 226–231, 2009. 18

[16] S. Katz, A. Ford, R. Moskowitz, B. Jackson, and M. Jaffe, "Studies of illness in the aged: The index of adl: a standardized measure of biological and psychosocial function," *JAMA*, vol. 185, no. 12, pp. 914–919, 1963. 18

[17] J. C. Millan-Calenti, J. Tubío, S. Pita-Fernández, I. González-Abraldes, T. Lorenzo, T. Fernández-Arruty, and A. Maseda, "Prevalence of functional disability in activities of daily living (adl), instrumental activities of daily living (iadl) and associated factors, as predictors of morbidity and mortality," *Archives of gerontology and geriatrics*, vol. 50, no. 3, pp. 306–310, 2010. 19, 68

[18] N. M. Gil, N. A. Hine, J. L. Arnott, J. Hanson, R. G. Curry, T. Amaral, and D. Osipovic, "Data visualisation and data mining technology for supporting care for older people," in *Proceedings of International ACM SIGACCESS Conference on Computers and accessibility*, ser. Assets '07, New York, USA, 2007, pp. 139–146. 19, 74

[19] H. Dodge, N. Mattek, D. Austin, T. Hayes, and J. Kaye, "In-home walking speeds and variability trajectories associated with mild cognitive impairment," *Neurology*, vol. 78, no. 24, pp. 1946–1952, 2012. 19

[20] S. Brownsell, S. Blackburn, and M. S. Hawley, "An evaluation of second and third generation telecare services in older people's housing," *Journal of Telemedicine and Telecare*, vol. 14, no. 1, pp. 8–12, 2008. 20

[21] "Gerhome project." accessed on 1/7/2015 at 15:00. [Online]. Available: http://gerhome.cstb.fr/en/ 21

[22] "Hermes project." accessed on 2/7/2015 at 15:00. [Online]. Available: http://www.fp7-hermes.eu/ 21

[23] S. Fleck, R. Loy, C. Vollrath, F. Walter, and W. Strasser, "Smartclassysurv-a smart camera network for distributed tracking and activity recognition and its application to assisted living," in *Proceeding of IEEE International Conference on Distributed Smart Cameras*, 2007, pp. 211–218. 21

[24] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Fall detection from human shape and motion history using video surveillance," in *Proceedings of International Conference on Advanced Information Networking and Applications Workshops (AINAW)*, vol. 2, 2007, pp. 875–880. 21

[25] B. Jansen and R. Deklerck, "Context aware inactivity recognition for visual fall detection," in *Proceedings of Pervasive Health Conference and Workshops*. IEEE, 2006, pp. 1–4. 21

[26] D. Chen, A. Bharucha, and H. Wactlar, "Intelligent video monitoring to improve safety of older persons," in *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, 2007, pp. 3814 –3817. 21

[27] G. E. Smith, A. M. Lunde, J. C. Hathaway, and K. S. Vickers, "Telehealth home monitoring of solitary persons with mild dementia," *American Jour-*

nal of Alzheimer's Disease and Other Dementias, vol. 22, no. 1, pp. 20–26, 2007. 21

[28] J. W. Davis and A. F. Bobick, "The representation and recognition of human movement using temporal templates," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1997, pp. 928–934. 22, 91, 92

[29] A. Albu, T. Beugeling, N. Virji-Babul, and C. Beach, "Analysis of irregularities in human actions with volumetric motion history images," in *Proceedings of IEEE Workshop on Motion and Video Computing (WMVC)*, 2007, pp. 16–16. 22

[30] D. Weinland, R. Ronfard, and E. Boyer, "Free viewpoint action recognition using motion history volumes," *Computer Vision and Image Understanding*, vol. 104, no. 2, pp. 249–257, 2006. 22, 122

[31] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2005, pp. 886–893 vol. 1. 22, 88, 99

[32] N. T. Nguyen, D. Q. Phung, S. Venkatesh, and H. Bui, "Learning and detecting activities from movement trajectories using the hierarchical hidden markov model," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2005, pp. 955–960 vol. 2. 22

[33] A. Gupta and L. S. Davis, "Objects in action: An approach for combining action understanding and object perception," in *Proceedings of IEEE Con-*

ference on Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–8. 22

[34] R. Messing, C. Pal, and H. Kautz, "Activity recognition using the velocity histories of tracked keypoints," in *Proceedings of IEEE Computer Society International Conference on Computer Vision (ICCV)*. Washington, DC, USA: IEEE Computer Society, 2009. 23, 70

[35] H. Cheng, Z. Liu, Y. Zhao, and G. Ye, "Real world activity summary for senior home monitoring," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, 2011, pp. 1–4. 23, 72

[36] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1290–1297. 23, 71

[37] G. J. Brostow and I. A. Essa, "Motion based decompositing of video," in *Proceedings of IEEE International Conference on Computer Vision*, vol. 1, 1999, pp. 8–13. 23

[38] M. Valstar, M. Pantic, and I. Patras, "Motion history for facial action detection in video," in *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, vol. 1, 2004, pp. 635–640 vol.1. 23

[39] A. N. Stein, "Occlusion boundaries: Low-level detection to high-level reasoning," Ph.D. dissertation, The Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, 2008. 23

[40] J. Fei and I. Pavlidis, "Thermistor at a distance: unobtrusive measurement of breathing," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 988–998, 2010. 24

[41] Z. Zhou, X. Chen, Y.-C. Chung, Z. He, T. X. Han, and J. M. Keller, "Activity analysis, summarization, and visualization for indoor human activity monitoring," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1489–1498, 2008. 24

[42] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1505–1518, 2003. 24

[43] M. Eldib, F. Deboeverie, W. Philips, and H. Aghajan, "Behavior analysis for elderly care using a network of low-resolution visual sensors," *Journal of Electronic Imaging*, vol. 25, no. 4, p. 041003, 2016. 24

[44] Q. Ni, A. B. García Hernando, and I. P. de la Cruz, "The elderlys independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development," *Sensors*, vol. 15, no. 5, pp. 11 312–11 362, 2015. 24

[45] S. Hargreaves, M. Hawley, A. Haywood, and P. Enderby, "Informing the design of lifestyle monitoring technology for the detection of health deterioration in long-term conditions: a qualitative study of people living with heart failure," *Journal of Medical Internet Research*, vol. 19, no. 6, June 2017. 25

[46] S. Pal and C. Abhayaratne, "Video-based activity level recognition for assisted living using motion features," in *Proceedings of the 9th International Conference on Distributed Smart Cameras*, ser. ICDSC '15. New York, NY, USA: ACM, 2015, pp. 62–67. 27

[47] N. Twomey, T. Diethe, M. Kull, H. Song, M. Camplani, S. Hannuna, X. Fafoutis, N. Zhu, P. Woznowski, and P. Flach, "The SPHERE chal-

lenge: Activity recognition with multimodal sensor data," *arXiv preprint arXiv:1603.00797*, 2016. 27

[48] L. Tao, T. Burghardt, M. Mirmehdi, D. Aldamen, A. Cooper, M. Camplani, S. Hannuna, A. Paiement, and I. Craddock, "Real-time estimation of physical activity intensity for daily living," in *Proceedings of the 2nd IET International Conference on Technologies for Active and Assisted Living (TechAAL 2016)*, 2016, pp. 11–16. 27

[49] M. M. Edwards, "The reliability and validity of self-report activities of daily living scales," *Canadian Journal of Occupational Therapy*, vol. 57, no. 5, pp. 273–278, 1990. 29

[50] L. Lorenzen-Huber, M. Boutain, L. J. Camp, K. Shankar, and K. H. Connelly, "Privacy, technology, and aging: A proposed framework," *Ageing International*, vol. 36, no. 2, pp. 232–252, 2011. 30

[51] C. L. Bentley, L. A. Powell, A. Orrell, and G. A. Mountain, "Addressing design and suitability barriers to telecare use: Has anything changed?" *Technology and Disability*, vol. 26, no. 4, pp. 221–235, 2014. 32

[52] A. Queirós, A. Silva, J. Alvarelhão, N. P. Rocha, and A. Teixeira, "Usability, accessibility and ambient-assisted living: a systematic literature review," *Universal Access in the Information Society*, vol. 14, no. 1, pp. 57–66, 2015. 32

[53] T. Jokela, N. Iivari, J. Matero, and M. Karukka, "The standard of user-centered design and the standard definition of usability: analyzing iso 13407 against iso 9241-11," in *Proceedings of ACM Latin American conference on Human-computer interaction*, 2003, pp. 53–60. 32

[54] D. A. Norman, "Cognitive engineering," *User centered system design*, pp. 31–61, 1986. 33

[55] A. Robinet, R. Picking, and V. Grout, "A framework for improving user experience in ambient assisted living," in *Proceedings of Fourth Collaborative Research Symposium on Security, E-learning, Internet and Networking (SEIN 2008), Glyndwr University, Wrexham*, 2008, pp. 111–121. 34

[56] Z. Callejas and R. López-Cózar, "Designing smart home interfaces for the elderly," *ACM SIGACCESS Accessibility and Computing*, no. 95, pp. 10–16, 2009. 34

[57] G. X. Ayala and J. P. Elder, "Qualitative methods to ensure acceptability of behavioral and social interventions to the target population," *Journal of public health dentistry*, vol. 71, no. s1, 2011. 34

[58] S. T. Londei, J. Rousseau, F. Ducharme, A. St-Arnaud, J. Meunier, J. Saint-Arnaud, and F. Giroux, "An intelligent videomonitoring system for fall detection at home: perceptions of elderly people," *Journal of Telemedicine and Telecare*, vol. 15, no. 8, pp. 383–390, 2009. 35

[59] L. Boise, K. Wild, N. Mattek, M. Ruhl, H. H. Dodge, and J. Kaye, "Willingness of older adults to share data and privacy concerns after exposure to unobtrusive in-home monitoring," *Gerontechnology : International Journal on the fundamental aspects of Technology to serve the ageing society*, vol. 11, no. 3, pp. 428–435, 2013. 35

[60] R. A. Krueger and M. A. Casey, *Focus groups: A practical guide for applied research*. Sage publications, 2014. 36

[61] J. Kitzinger, "Qualitative research. introducing focus groups." *BMJ: British Medical Journal*, vol. 311, no. 7000, p. 299, 1995. 36

[62] J. Smithson, "Using and analysing focus groups: limitations and possibilities," *International Journal of Social Research Methodology*, vol. 3, no. 2, pp. 103–119, 2000. 36

[63] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006. 40

[64] R. B. Schafer and J. L. Tait, *A guide for understanding attitudes and attitude change*. Cooperative Extension Service, Iowa State University, 1981. 40

[65] H. Storf, T. Kleinberger, M. Becker, M. Schmitt, F. Bomarius, and S. Prueckner, "An event-driven approach to activity recognition in ambient assisted living," in *Proceedings of European Conference on Ambient Intelligence*. Springer, 2009, pp. 123–132. 52

[66] S. Brownsell, D. Bradley, F. Cardinauxm, and M. Hawley, "Developing a systems and informatics based approach to lifestyle monitoring within ehealth: part i-technology and data management," in *Healthcare Informatics, Imaging and Systems Biology (HISB), 2011 First IEEE International Conference on*. IEEE, 2011, pp. 264–271. 53

[67] S. Brownsell, D. Bradley, F. Cardinaux, and M. Hawley, "Developing a systems and informatics based approach to lifestyle monitoring within ehealth: part ii-analysis & interpretation," in *Healthcare Informatics, Imaging and Systems Biology (HISB), 2011 First IEEE International Conference on*. IEEE, 2011, pp. 213–220. 53

[68] ——, "Developing a Systems and Informatics based approach to lifestyle monitoring within ehealth: Part I - Technology and Data Management,"

in *Proceedings of IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology (HISB)*, 2011, pp. 264–271. 59

[69] D. Curone, G. M. Bertolotti, A. Cristiani, E. L. Secco, and G. Magenes, "A real-time and self-calibrating algorithm based on triaxial accelerometer signals for the detection of human posture and activity," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 4, pp. 1098–1105, 2010. 69, 75

[70] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: a local svm approach," in *Proceedings of IEEE International Conference on Pattern Recognition, (ICPR)*, vol. 3, 2004, pp. 32–36. 70

[71] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2005, pp. 1395–1402. 70

[72] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2247–2253, 2007. 70

[73] "CAVIAR dataset by EC Funded CAVIAR Project/IST 2001 37540," accessed on 1/7/2015 at 15:00. [Online]. Available: http://homepages.inf.ed.ac.uk/rbf/CAVIAR/ 70

[74] M. Tenorth, J. Bandouch, and M. Beetz, "The TUM kitchen data set of everyday manipulation activities for motion tracking and action recognition," in *Proceedings of IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2009, pp. 1089–1096. 71

[75] M. Rohrbach, S. Amin, M. Andriluka, and B. Schiele, "A database for fine grained activity detection of cooking activities," in *Proceedings of the IEEE*

Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1194–1201. 72

[76] N. Hine, A. Judson, S. Ashraf, J. Arnott, A. Sixsmith, S. Brown, and P. Garner, "Modelling the behaviour of elderly people as a means of monitoring well being," in *Proceedings of International Conference on User Modelling*, ser. UM'05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 241–250. 73

[77] B. E. Ainsworth, W. L. Haskell, S. D. Herrmann, N. Meckes, D. R. Bassett Jr, C. Tudor-Locke, J. L. Greer, J. Vezina, M. C. Whitt-Glover, and A. S. Leon, "2011 compendium of physical activities: a second update of codes and met values." *Medicine and science in sports and exercise*, vol. 43, no. 8, pp. 1575–1581, 2011. 82

[78] Y. Chin and C. J. Tsai, "Dense true motion field compensation for video coding," in *2013 IEEE International Conference on Image Processing*, Sept 2013, pp. 1958–1961. 88

[79] N. Dalal, "Finding people in images and videos," Ph.D. dissertation, Institut National Polytechnique de Grenoble-INPG, 2006. 88, 93, 99

[80] B. Jähne, H. Haussecker, and P. Geissler, *Handbook of computer vision and applications.* Citeseer, 1999, vol. 2. 90, 95

[81] A. Mitiche and J. K. Aggarwal, *Computer vision analysis of image motion by variational methods.* Springer Science & Business Media, 2013, vol. 10. 90

[82] B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981. 91

[83] T. Brox and J. Malik, "Large displacement optical flow: descriptor matching in variational motion estimation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 3, pp. 500–513, 2011. 91

[84] D. Weinland, R. Ronfard, and E. Boyer, "A survey of vision-based methods for action representation, segmentation and recognition," *Computer vision and image understanding*, vol. 115, no. 2, pp. 224–241, 2011. 91

[85] L. A. Schwarz, A. Mkhitaryan, D. Mateus, and N. Navab, "Human skeleton tracking from depth data using geodesic distances and optical flow," *Image and Vision Computing*, vol. 30, no. 3, pp. 217–226, 2012. 91

[86] H. S. Parekh, D. G. Thakore, and U. K. Jaliya, "A survey on object detection and tracking methods," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 2, pp. 2970–2979, 2014. 91

[87] M. A. Ahad, "Motion history image," in *Motion History Images for Action Recognition and Understanding.* Springer, 2013, pp. 31–76. 92

[88] G. R. Bradski and J. W. Davis, "Motion segmentation and pose recognition with motion history gradients," *Machine Vision and Applications*, vol. 13, no. 3, pp. 174–184, 2002. 92

[89] J. Davis and G. Bradski, "Real-time motion template gradients using intel cvlib," in *Proceedings of IEEE International Conference on Computer Vision (ICCV) Workshop on Frame-rate Vision*, 1999, pp. 1–20. 92

[90] M. B. Kaniche and F. Brmond, "Recognizing gestures by learning local motion signatures of hog descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2247–2258, 2012. 93

[91] M. B. Kaaniche and F. Bremond, "Tracking hog descriptors for gesture recognition," in *Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2009, pp. 140–145. 93

[92] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *IJCAI*, vol. 81, 1981, pp. 674–679. 96

[93] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/kanade meets horn/schunck: Combining local and global optic flow methods," *International Journal of Computer Vision*, vol. 61, no. 3, pp. 211–231, 2005. 96

[94] J. Y. Bouguet, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," *Intel Corporation*, vol. 5, no. 1-10, p. 4, 2001. 96

[95] D. D. Doyle, A. L. Jennings, and J. T. Black, "Optical flow background estimation for real-time pan/tilt camera object tracking," *Measurement*, vol. 48, pp. 195–207, 2014. 97

[96] A. Plyer, G. Le Besnerais, and F. Champagnat, "Massively parallel lucas kanade optical flow for real-time video processing applications." *Journal of Real-Time Image Processing*, vol. 11, no. 4, pp. 713–730, 2016. 97

[97] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, "Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1932–1939. 99

[98] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented his-

tograms of flow and appearance," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2006, pp. 428–441. 99

[99] J. Zhang and S. Gong, "Action categorization with modified hidden conditional random field," *Pattern Recognition*, vol. 43, no. 1, pp. 197–203, 2010. 100

[100] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962. 101

[101] G. Hinton, L. Deng, D. Yu, A. rahman Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. S. G. Dahl, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012. 102

[102] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of Clinical Epidemiology*, vol. 49, no. 11, pp. 1225 – 1231, 1996. 102

[103] V. Vapnik, *Statistical Learning Theory*, ser. Population Studies. John Wiley and Sons, 1998, no. 207. 102

[104] A. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes," *Advances in Neural Information Processing Systems*, vol. 14, p. 841, 2002. 102

[105] K. Wang, X. Wang, L. Lin, M. Wang, and W. Zuo, "3d human activity recognition with reconfigurable convolutional neural networks," in *Proceedings of ACM International Conference on Multimedia*, ser. MM 2014, New York, NY, USA, 2014, pp. 97–106. 102

[106] S. Oniga and J. Suto, "Human activity recognition using neural networks," in *Proceedings of International Carpathian Control Conference (ICCC)*, 2014, pp. 403–406. 102

[107] B. G. Kermani, S. S. Schiffman, and H. T. Nagle, "Performance of the levenberg–marquardt neural network training method in electronic nose applications," *Sensors and Actuators B: Chemical*, vol. 110, no. 1, pp. 13–22, 2005. 103

[108] I. G. Karybali, E. Z. Psarakis, K. Berberidis, and G. D. Evangelidis, "An efficient spatial domain technique for subpixel image registration," *Signal Processing: Image Communication*, vol. 23, no. 9, pp. 711–724, 2008. 120

[109] V. Argyriou and T. Vlachos, "A study of sub-pixel motion estimation using phase correlation," in *Proceedings of British Machine Vision Conference (BMVC)*, 2006, pp. 40.1–40.10. 120

[110] F. Zhang, W. Jiang, F. Autrusseau, and W. Lin, "Exploring v1 by modeling the perceptual quality of images," *Journal of vision*, vol. 14, no. 1, pp. 26–26, 2014. 121, 122

[111] T. Lindeberg, "Invariance of visual operations at the level of receptive fields," *BMC Neuroscience*, vol. 14, no. 1, p. P242, 2013. 121

[112] L. Shen, Z. Liu, X. Zhang, W. Zhao, and Z. Zhang, "An effective cu size decision method for hevc encoders," *IEEE Transactions on Multimedia*, vol. 15, no. 2, pp. 465–470, 2013. 121, 124

[113] G. J. Sullivan, J. M. Boyce, Y. Chen, J.-R. Ohm, C. A. Segall, and A. Vetro, "Standardized extensions of high efficiency video coding (hevc)," *IEEE Journal of selected topics in Signal Processing*, vol. 7, no. 6, pp. 1001–1016, 2013. 121

[114] K. Xue, X. Wang, G. Ma, H. Wang, and D. Nam, "A video saliency detection method based on spatial and motion information," in *International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 412–416. 121

[115] M. B. Patel, D. R. Kshirsagar, and V. Nitnaware, "Review and comparative study of motion estimation techniques to reduce complexity in video compression," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 2, no. 8, pp. 3574–3584, 2013. 121

[116] K. Müller, A. Maier, C. Schwemmer, G. Lauritsch, S. De Buck, J. Wielandts, J. Hornegger, and R. Fahrig, "Image artefact propagation in motion estimation and reconstruction in interventional cardiac c-arm ct," *Physics in medicine and biology*, vol. 59, no. 12, p. 3121, 2014. 121

[117] J. D. Power, A. Mitra, T. O. Laumann, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Methods to detect, characterize, and remove motion artifact in resting state fmri," *Neuroimage*, vol. 84, pp. 320–341, 2014. 121

[118] I. Kotsia and V. Argyriou, "Action spotting exploiting the frequency domain," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. IEEE, 2011, pp. 43–48. 122

[119] H. Imtiaz, U. Mahbub, G. Schaefer, S. Y. Zhu, and M. A. R. Ahad, "Human action recognition based on spectral domain features," *Procedia Computer Science*, vol. 60, pp. 430–437, 2015. 122

[120] A. Tran, J. Guan, T. Pilantanakitti, and P. Cohen, "Action recognition in the frequency domain," *arXiv preprint arXiv:1409.0908*, 2014. 122

[121] E. Gladilin and R. Eils, "On the role of spatial phase and phase correlation

in vision, illusion, and cognition," *Frontiers in computational neuroscience*, vol. 9, 2015. 122, 124

[122] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE transactions on medical imaging*, vol. 32, no. 7, pp. 1153–1190, 2013. 124

[123] S. Pal, T. Feng, and C. Abhayaratne, "Real-time recognition of activity levels for ambient assisted living," in *Proceedings of the 5th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*. IEEE, 2015, pp. 485–488. 171

Appendices

Appendix A

Different Activities of the SADL dataset

Images of the different activities of the SADL dataset are presented here. All the 9 identified activities as listed in Table 4.5 under the two illumination conditions from each of the camera angles. As mentioned in Chapter 4, both the camera angles are synchronized for the dataset.

(a) Camera 1

(b) Camera 2

Figure A.1: 'Use Fridge' activity from both Camera Angles with indoor light switched on

(a) Camera 1

(b) Camera 2

Figure A.2: 'Use Over (Plate In)' activity from both Camera Angles with indoor light switched on

(a) Camera 1

(b) Camera 2

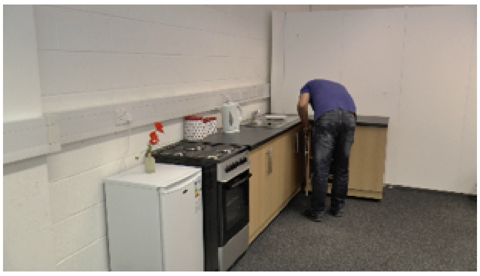
Figure A.3: 'Use Over (Plate Out)' activity from both Camera Angles with indoor light switched on

(a) Camera 1

(b) Camera 2

Figure A.4: 'Pour Water' activity from both Camera Angles with indoor light switched on

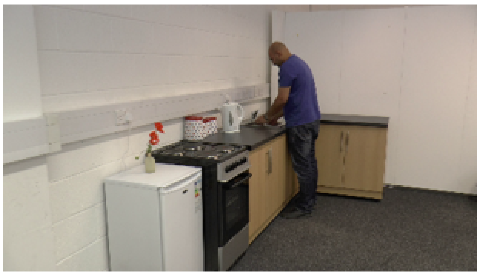
(a) Camera 1



(b) Camera 2

Figure A.5: ‘Use Cupboard’ activity from both Camera Angles with indoor light switched on

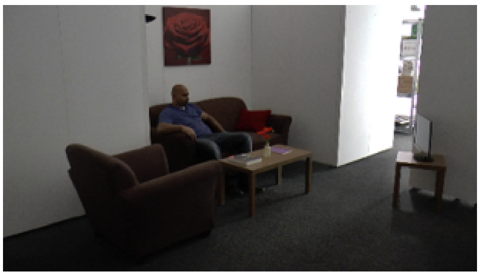
(a) Camera 1



(b) Camera 2

Figure A.6: ‘Wash Plate’ activity from both Camera Angles with indoor light switched on

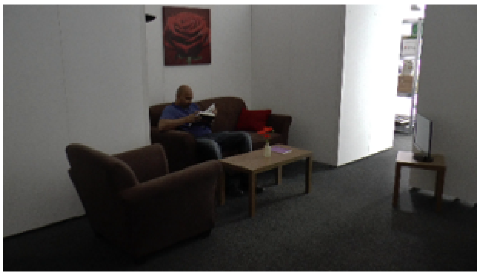
(a) Camera 1



(b) Camera 2

Figure A.7: ‘Watch Television’ activity from both Camera Angles with indoor light switched off

(a) Camera 1



(b) Camera 2

Figure A.8: 'Reading Book' activity from both Camera Angles with indoor light switched off

(a) Camera 1

(b) Camera 2

Figure A.9: 'Sitting Idle' activity from both Camera Angles with indoor light switched off

(a) Camera 1

(b) Camera 2

Figure A.10: 'Walking' activity in the Kitchen from both Camera Angles with indoor light switched on

(a) Camera 1

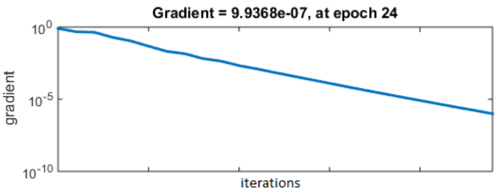
(b) Camera 2

Figure A.11: ‘Walking’ activity in the Living room from both Camera Angles with indoor light switched off

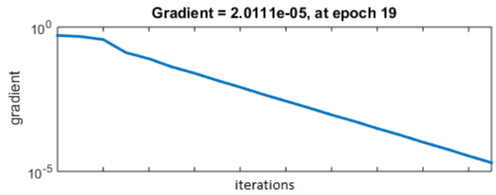
Appendix B

Training State Results of the Trained Neural Network Model for various Experiments

The training state results for all the experiments are presented in this appendix. Training state results show the number of iterations (epochs) needed for each of the neural network model to reach the minimum gradient. The minimum gradient and the iteration at which the training was stopped is mentioned along with the each of the figures.

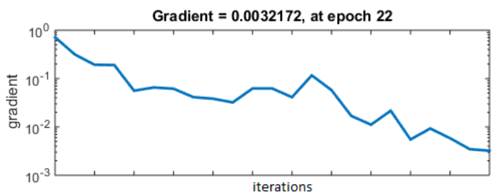


(a) Camera 1

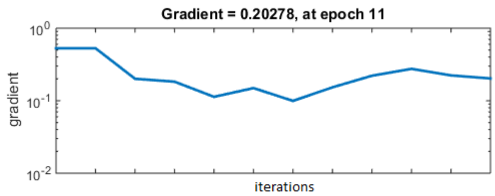


(b) Camera 2

Figure B.1: Training state results for the motion-HOG feature under single camera setup

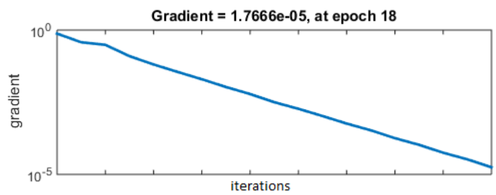


(a) Camera 1

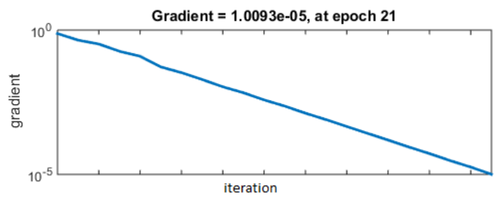


(b) Camera 2

Figure B.2: Training state results for the motion-moment feature under single camera setup

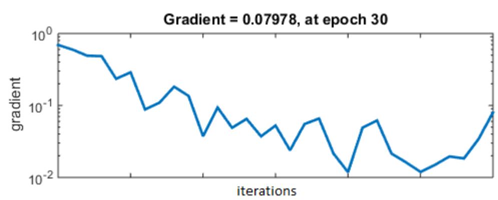


(a) Camera 1

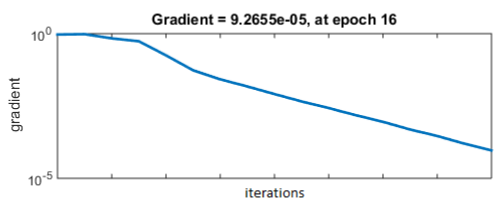


(b) Camera 2

Figure B.3: Training state results for all the three features under single camera setup

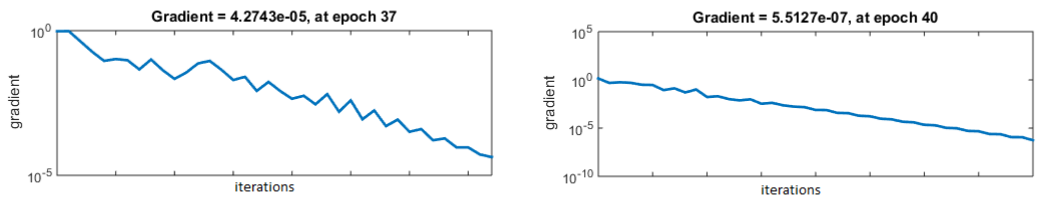


(a) Baseline



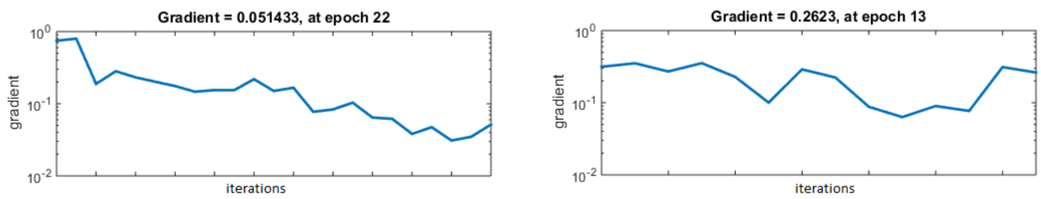
(b) motion-HOG feature

Figure B.4: Training state results for the baseline experiment and motion-HOG feature under dual camera setup



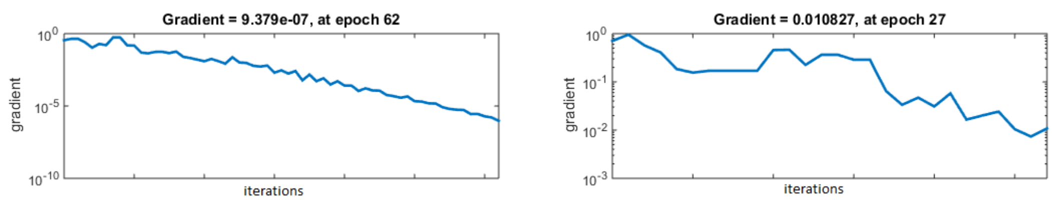
(a) motion-moment feature (b) all 3 features

Figure B.5: Training state results for the motion-moment feature and all three features fused together under dual camera setup



(a) $-\pi$ to π (b) $-\pi/2$ to $\pi/2$

Figure B.6: Training state results for the phase-based features for real axis ranges



(a) $-\pi/4$ to $\pi/4$ (b) $-\pi/8$ to $\pi/8$

Figure B.7: Training state results for the phase-based features for real axis ranges

Appendix C

Additional Results for the Real Time Experiment

For a continuous monitoring system, estimation of activity levels should be done at real-time. To test the initial assumptions of using motion vectors and motion density as features for activity level estimation, the experiments are further simulated on the low computation processing Beagle board. The figure and the description of the beagle-board are presented shown in Appendix C.

The performance of the classification is the same as the baseline results. However, in real time applications, along with accuracy, computation time is also impor-

Table C.1: Detection Time on Beagle board

	Average Detection Time (in seconds)
Subject 1	27.89
Subject 2	26.38

tant. Since the feature extraction and the estimation task was carried out using a low computation power processor, the time to process and extract features from memory intensive data like images is critical. This experiment was carried out to test the practicality of an activity level estimation system for a continuous monitoring solution [123]. The time shown in TABLE C.1, is the average time required to predict an activity level.

Beagle Board is an open source, low computation power, low cost hardware designed by Texas Instruments. The BeagleBoard used for the experiments is the Rev C3 version. It has a ARM Cortex-A8 processor (OMAP3530) with an external microSD card of 4GB. The board has been booted up with the LINUX distribution.

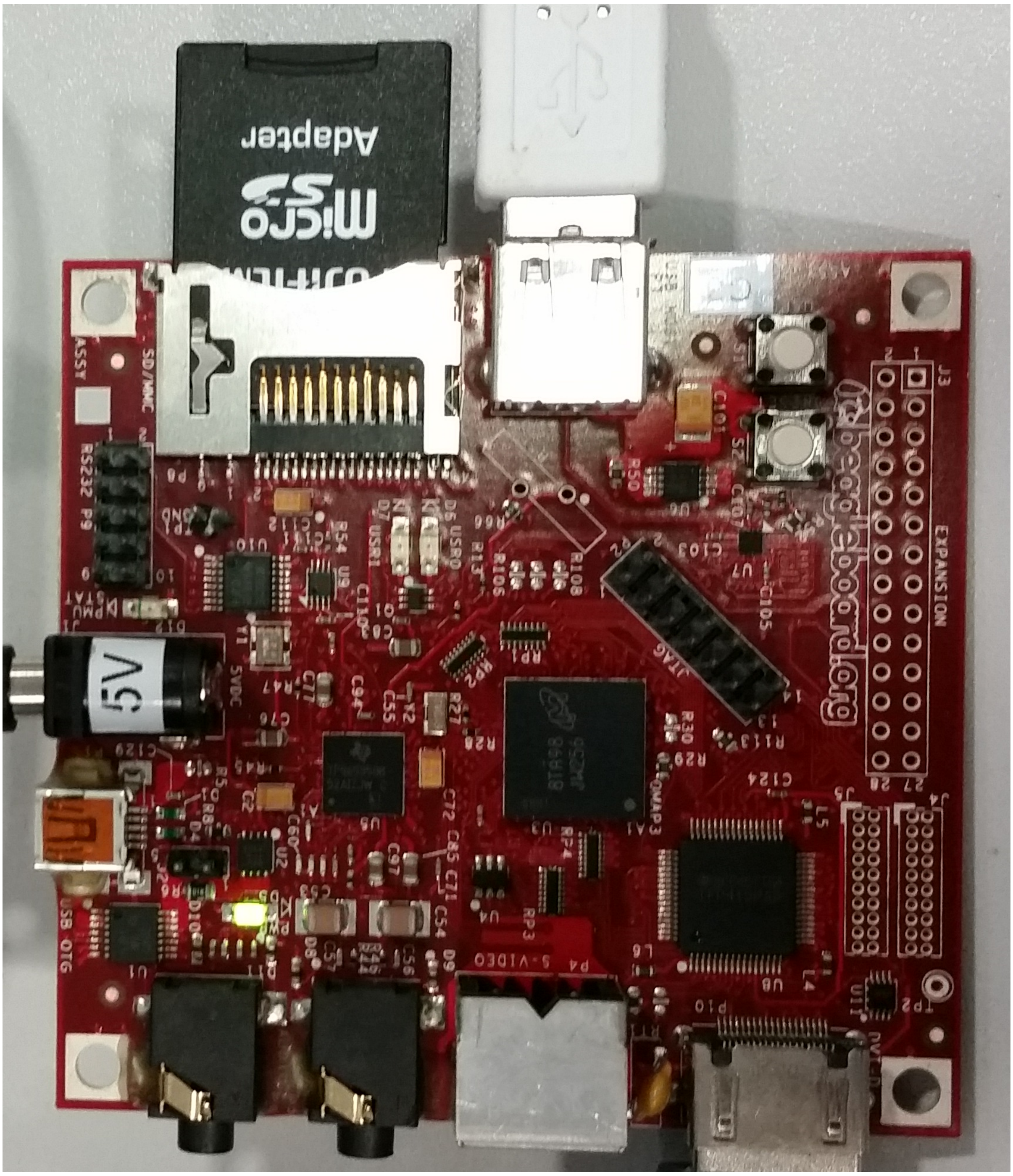


Figure C.1: Beagle-board

Appendix D

SADL Dataset Consent Forms

The actors in the dataset were colleagues within the research group and each of them had voluntarily participated in the activity. Following are the signed consent forms for participation.

The Consent forms have been removed for the electronic version.

Appendix E

Ethical Clearance

Ethics Application The following application was used to obtain the Ethical Clearance from The University of Sheffield to conduct the Focus Groups. The application was reviewed by the Ethics reviewers of the Department of Electronic and Electrical Engineering and the School of Health and Related Research within The University of Sheffield.

University Research Ethics Application Form for Staff and PGRs

This form has been approved by the University Research Ethics Committee (UREC)

Date:	
Name of applicant:	Mr. Sandipan Pal
Research project title:	Adaptive Lifestyle Monitoring

Complete this form if you are a **member of staff or a postgraduate research student** who plans to undertake a research project which requires ethics approval via the University Ethics Review Procedure. ☑

or

Complete this form if you plan to submit a **'generic' research ethics application (i.e. an application** that will cover several sufficiently similar research projects). Information on the 'generic' route is at: www.sheffield.ac.uk/ris/other/gov-ethics/ethicspolicy/approval-procedure/review-procedure/generic-research-projects ☐

If you are an undergraduate or a postgraduate-taught student, this is the wrong form.

PLEASE NOTE THAT YOUR DEPARTMENT MAY USE A VARIATION OF THIS FORM: PLEASE CHECK WITH THE ETHICS ADMINISTRATOR IN YOUR DEPARTMENT

This form should be accompanied, where appropriate, by all Information Sheets/Covering Letters/Written Scripts which you propose to use to inform the prospective participants about the proposed research, and/or by a Consent Form where you need to use one.

Further guidance on how to apply is at: www.shef.ac.uk/ris/other/gov-ethics/ethicspolicy/approval-procedure/review-procedure

Guidance on the possible routes for obtaining ethics approval (i.e. on the University Ethics Review Procedure, the NHS procedure and the Social Care Research Ethics Committee, and the Alternative procedure) is at: www.shef.ac.uk/ris/other/gov-ethics/ethicspolicy/approval-procedure/ethics-approval

Once you have completed this research ethics application form in full, and other documents where appropriate, check that your name, the title of your research project and the date is contained in the footer of each page and email it to the Ethics Administrator of your academic department. Please note that the original signed and dated version of 'Part B' of the application form should also be provided to the Ethics Administrator in hard copy. Ethics Administrators are listed at:
www.shef.ac.uk/polopoly_fs/1.99105!/file/Ethics-Administrators.pdf

I confirm that I have read the current version of the University of Sheffield 'Ethics Policy Governing Research Involving Human Participants, Personal Data and Human Tissue', as shown on the University's research ethics website at: www.shef.ac.uk/ris/other/gov-ethics/ethicspolicy ☑

Part A

A1. Title of Research Project: Adaptive Lifestyle Monitoring

A2. Contact person (normally the Principal Investigator, in the case of staff-led research projects, or the student in the case of supervised-postgraduate researcher projects):

Title: Mr.
Post: PGR student
Email: s.pal@sheffield.ac.uk

Name: Sandipan Pal
Department: : Department of Electronic and Electrical Engineering
Telephone:

A2.1. Is this a postgraduate researcher project? If yes, please provide the Supervisor's contact details:

Title: Dr.
Post: Lecturer
Email: c.abhayaratne@sheffield.ac.uk

Name: Charith Abhayaratne
Department: Department of Electronic and Electrical Engineering
Telephone: (+44) (0)114 222 5893

A2.2. Other key investigators/co-applicants (within/outside University), where applicable. Please list all (add more if necessary):

Title: Prof.
Post: Professor
Email: mark.hawley@sheffield.ac.uk

Name: Mark Hawley
Department: School of Health and Related Research
Telephone: (+44) (0)114 222 0682

Title:
Post:
Email:

Name:
Department:
Telephone:

A3. Proposed Project Duration:

Start date: 1/10/2011

End date: 30/9/2015

A4. Mark 'X' in one or more of the following boxes if your research:

	involves adults with mental incapacity or mental illness
	involves prisoners or others in custodial care (e.g. young offenders)
	involves children or young people aged under 18 years
	involves using samples of human biological material collected before for another purpose
	involves taking new samples of human biological material (e.g. blood, tissue) *
	involves testing a medicinal product *
	involves taking new samples of human biological material (e.g. blood, tissue) *
	involves additional radiation above that required for clinical care *
	involves investigating a medical device *
	is social care research
	is ESRC funded

* If you have marked boxes marked * then you also need to obtain confirmation that appropriate University insurance is in place. The procedure for doing so is entirely by email. Please send an email addressed to insurance@shef.ac.uk and request a copy of the 'Clinical Trial Insurance Application Form'.

It is recommended that you familiarise yourself with the University's Ethics Policy Governing Research Involving Human Participants, Personal Data and Human Tissue before completing the following questions. Please note that if you provide sufficient information about the research (what you intend to do, how it will be carried out and how you intend to minimise any risks), this will help the ethics reviewers to make an informed judgement quickly without having to ask for further details.

A5. Briefly summarise:

i. The project's aims and objectives:
(this must be in language comprehensible to a lay person)

Aim –
The main aim of the project is to address the technological challenges involved in the design and development of a video-based lifestyle monitoring system.

Objectives –
1. Identification of the socio-technical relevance of a video-based lifestyle monitoring system – a qualitative study to explore the relevance and acceptance of a camera-based technology.
2. To study the various computer vision techniques for feature extraction and tracking
3. To design an activity level detector using motion patterns within video sequences
4. To fuse the activity level information with postural features for event detection.
5. To propose a complete framework for a video-based monitoring system for assisted living through monitoring of the individual along with the ambient parameter.

In this research we would like to explore the use of camera instead of the conventional sensors for a monitoring technology.
This study is intended to attain the objective number 1 of the project.
The alarming increase in the elderly population in the next 50 years envisages an increasing need of technology to support the daily living of elderly people. Research has shown that there might be a link between the health of a person and the daily activities that is being undertaken. The technologies are deployed to monitor the activities and draw conclusions to foresee a potential decline in the health condition. Most of the present lifestyle monitoring technologies is typically sensor-based solutions (like smoke detectors, infra-red sensors), where a number of embedded or body-worn sensors are deployed or connected over a network to monitor activities of an individual. In this research we would want to explore the use of camera for monitoring the activities.
A lifestyle monitoring system should be a continuous collection of non-intrusive data which is representative of ones well-being. Visual Data is contextually rich and can provide with a lot of information about ones' daily living. However being constantly recorded by a camera is extremely intrusive. This study is aimed at achieving the objective number 1 of the project that is to understand the relevance and potential privacy concerns of the users regarding a video-based monitoring system.
We intend to organize focus group discussion to identify the possible concerns of the users regarding video-based monitoring systems. Focus groups are group discussions involving 6-8 people to discuss about a certain topic.

ii. **The project's methodology:**
(this must be in language comprehensible to a lay person)

Focus groups are a qualitative research technique to understand the opinions, concerns of the potential users of the system. A monitoring technology has two stake-holders, the carers and the users. In our focus groups we will address the concerns of the latter group. Though the carers have expressed their displeasure in only addressing the concerns of the users in [1] while designing such technology; most of the technologies used or discussed about in that research was sensor-based. The existing acceptability studies done so far suggest that users would accept video-based system in principle with the condition that they feel that it would make a real difference to their well-being. However, the gap between the acceptability in principle and the consent to adopt the technology can only be reduced by developing the technology together with the study of user needs [2]. Usually 3 to 4 groups with selected for a focus group study [3].

Study participants
Anyone who falls within the age category of the groups is eligible to participate in the discussion. No specific expertise, technical knowledge or medical condition is required for participating in this study.

Inclusion criteria –

Any individual who falls within the particular age bracket [above 70years, 45-70years] and are healthy can participate in the discussion.

Participants who have a little understanding of basic technology and enjoy living independently would be particularly welcomed.

The participants would ned to be physically mobile and relatively healthy in order to commute to the location of where the focus group would be held. The venue for the focus group would definitely have the 'ease of access' option.

Adequate English language skills in order to understand and follow the instructions.

Must consent to being audio-recorded as a part of the focus group participation.

Living in the community.

Exclusion criteria –

A potential participant who meets any of the following criteria will be excluded from participation in this study:

Any severe medical condition that may affect the understanding and participation process.

Severe sensory deficits (specially speech and hearing)

Visual impairments (that cannot be corrected with glasses or contact lenses to within normal or near normal limits)

[The participants would be allowed to bring in their carers if they wish to. The carers can attend to the needs of the participants during the study. The researchers need to be informed before if the participants are bringing in carers. However the carer would not be allowed to make any contribution to the focus group discussions.]
The participants are expected to be involved only for about 1.5 hours to 2 hours (excluding time for travelling) for one day.

Location of the focus group meeting –
The Focus groups would be held within the newly setup HOMELAB of the University. The address for the HOMELAB is –

The Innovation Centre,
217 Portobello Street,
Sheffield S1 4DP

If the HOMELAB is unavailable, then a meeting room within the University will be booked.
[Please note that the venue for the group discussion would definitely have the *'ease of access'* option]

The **date** and the **time** for the discussion is negotiable based on the availability of the participants and the venue.
The time would be scheduled within the working hours of a normal working day.

The focus group discussion would ideally last for about 1.5 hours to 2 hours and would be broken up into 4 stages

PLAN for the focus groups
Stage I : Introduction

1. Welcome

2. A brief introductory presentation about the study

3. Individual introductions

Stage II : The opening discussion

1. Presentation on Video-based monitoring Technology and its advantages (demos of gesture recognition or other computer vision based examples would be shown. The demos are intended to gain confidence among the participants about the recent advancements of camera-based technologies in different domains.)

2. Discussion

Stage III : Demo and discussion

1. A brief demonstration of a video-based system (video/animation based)

2. Discussion

    a. What sort of activities/events should the technology identify for their profile?

    b. Where and how would they like the video information to be stored?

    c. To whom should this information be disclosed to? Would they like to control it?

Stage IV : Concluding remarks

Data Analysis
The data generated from the focus groups would be qualitatively analysed on two broad themes such as –

1. General attitude towards video-based monitoring system
2. Privacy concerns and acceptability

More themes might evolve during the analysis of the data

References-
[1] H. Thompson and S. Thielke, "How do health care providers perceive technologies for monitoring older adults?," in *Engineering in Medicine and Biology Society, 2009, EMBC 2009. Annual*

International Conference of IEEE, pp. 4315-4318, 2009.
[2] F. Cardinaux, D. Bhowmik, C. Abhayaratne and M. S. Hawley, "Video based technology for ambient assisted living: A review of the literature," *Journal of Ambient Intelligence and Smart Environments,* vol. 3, no. 3, pp. 253-269
[3] R. A. Krueger, *Focus groups: A practical guide for applied research.* Sage, 2009.

A6. What is the potential for physical and/or psychological harm/distress to participants?

There should none or very low risk of physical or psychological harm/ distress to the participants.
The focus groups and the discussions would be non-intrusive and would focus on how much a video based monitoring system would be acceptable.
Proper University identification will be shown while approaching the voluntary organizations or any individuals.
The focus group would take place during the working hours of a normal working weekday.

A7. Does your research raise any issues of personal safety for you or other researchers involved in the project? (especially if taking place outside working hours or off University premises)

No.
The Focus groups would be held within the University premises during the normal working hours of a normal working weekday.
A researcher may approach the voluntary organizations alone to reach the information booklet or speak about the proposed study. During such events, the researcher would notify his supervisors.
All the personal records of the participants would be kept confidential with the researchers.
During recruitment and also before participation, it would be explained to the participants that the group discussions are not entirely confidential. It would be explained that this is more of the discussion forum where they can discuss about their feelings on a particular topic.
More details about this is there on the Information Booklet.

If yes, explain how these issues will be managed.

A8. How will the potential participants in the project be:

i. Identified?

The study intends to have 3 focus groups with 6 to 10 people in each group. The participants will be identified based on their age and the participation would be completely voluntary. They can opt out of the discussion whenever they wish to.

Focus group No.	No. of participants	Age group	Purpose
1	6-10	Above 70 years	To identify the potential privacy concerns and their reactions to a video-based technology. To evaluate the reliability and acceptance of a camera-based monitoring technology.
2	6-10	Above 70 years	
3	6-10	Between 45 to 70 years	

Group 1 and 2 are identified as potential users of a video-based system. Group 3 has a different age range because they would be the future users of such a monitoring system. As technology has evolved rapidly over the last two decades, the perception towards a new technology of individuals between the age group of 45 and 70 is also supposed to be different.

ii. Approached?

The local voluntary organizations would be approached to recruit the participants of the focus groups. There will be provided with several copies of the Information booklets that they can distribute among their members. They would also be given a sample of the Consent form which the prospective participants may have a look at.
Participation in the study will be voluntary.
For the third group [age category between 45-70 years], volunteers from within the University may be asked to participate.
Potential participants who express interest in participation would be given an Information booklet along with the contact information of the researchers for any queries. The participants will be given a minimum of 24 hours to decide whether they would like to participate in the discussion. The researcher would try to gain confidence among the prospective participants and answer their queries and concerns about participation.
After the participants agree to participate in the discussion, if they want to withdraw, they would need to inform the researchers at least 48 hours before the scheduled date.
Before participation, all research participants would be asked to sign a 'Consent form'.

iii. Recruited?

Research participants would be recruited through local voluntary organizations. The participants who fall within the above mentioned age category can participate in the study. No specific expertise, technical knowledge or medical condition is required for participating in this study.
However since the proceeding of the discussion would be audio-recorded, and later transcribed, individuals with severe speech impairment would find it difficult to participate in the discussion.
The researchers would approach the voluntary organizations and set up a meeting with the administrative staff of the organization. During the meeting, he would explain about the study and also hand them over the information booklets to be distributed among the members. When approaching the organizations, the researcher would always carry the University ID.
The researcher would keep in touch with the organization and as soon as the number of participants is reached for one of the groups, a date based on the availability of the participants would be finalized.
This process would be continued till the proposed number of focus groups is conducted.

Carers involvement is subjected to pre-notification and the choice of the participants.

For group 3 [age bracket between 45 and 70 years], emails can be sent within the University network, to ask for participants.

A9. Will informed consent be obtained from the participants?

Yes ☑ No ☐

If informed consent or consent is NOT to be obtained please explain why. Further guidance is at: www.shef.ac.uk/ris/other/gov-ethics/ethicspolicy/policy-notes/consent

A9.1. This question is only applicable if you are planning to obtain informed consent:
How do you plan to obtain informed consent? (i.e. the proposed process?):

All participants will be provided with a copy of Information booklet and Consent Form prior to consent for participation being requested. The participants would be given at least 24 hours to think about whether they would like to participate in the discussion or not. They can also contact the researchers if they want to ask any questions about the study. After the participants have agreed, a suitable date would be fixed based on the availability of the participants.
The study and the questions would be explained to the participants prior to the study. The researcher will respond to all the queries of the participants regarding the study. The participation of all participants would be voluntary.
Before participation, the participants would have to sign a Consent Form. All the Consent Forms would be kept securely by the researchers for their record.
Details of this are explained in the Information booklet and consent form.

Remember to attach your consent form and information sheet (where appropriate)

A10. What measures will be put in place to ensure confidentiality of personal data, where appropriate?

All participants would sign a 'Consent form' that would confirm that their participation is voluntary and the data collected would only be used for research purpose. It has been clearly mentioned as to how the data will be collected and where it would be kept.
In the Information booklet, it is clearly mentioned that all personal details would be kept strictly confidential. Since focus groups are like group discussions, fellow participants might get to know about the views of a particular participant during the meeting. However no audio or transcribed files would be made available to them.
All the data will be stored securely in a password protected University computer. Back-ups might be kept in Departmental server for security. The identity of the participants will be encrypted using a number system (code name). After the audio recordings are transferred onto a university computer, the memory of the recorder would be formatted.
The identity of the individuals will not be disclosed in any report or publication. If any quotation is used, then the participant's identity will be made anonymous.
The notes taken during the discussion, the audio files, and the transcribed files will be available only for the researchers (PhD Student and supervisors). The files can be used only for research purposes.
After the conclusion of the project, the files would be further kept in the University archives for a period of 3 years and only the researcher associated with this project would have access to them. Following the 3 year period, the data would be deleted from the University servers.
The participants need to agree on this as mentioned in the Information booklet and the Consent Form.

A11. Will financial/in kind payments (other than reasonable expenses and compensation for time) be offered to participants? (Indicate how much and on what basis this has been decided)

No.
There would be light refreshments (tea, coffee, biscuits) during the discussions.
Reimbursement of the travel costs of the participants to the University is negotiable.

A12. Will the research involve the production of recorded media such as audio and/or video recordings?

YES ✓ NO ☐

A12.1. This question is only applicable if you are planning to produce recorded media:
How will you ensure that there is a clear agreement with participants as to how these recorded media may be stored, used and (if appropriate) destroyed?

Audio recordings.
The proceedings of the focus groups will be recorded. Permission will be sought before any recordings through the 'Consent form'. All information will be kept strictly confidential by the researchers. The recordings will be stored securely at the university and only will be used for the purposes of the research. After the project ends, the recordings will be stored in a secure, locked university archive space.

Guidance on a range of ethical issues, including safety and well-being, consent and anonymity, confidentiality and data protection are available at: www.shef.ac.uk/ris/other/gov-ethics/ethicspolicy/policy-notes

University Research Ethics Application Form - Part B - The Signed Declaration

Title of Research Project:

Adaptive Lifestyle Monitoring

I confirm my responsibility to deliver the research project in accordance with the University of Sheffield's policies and procedures, which include the University's '*Financial Regulations*', '*Good Research Practice Standards'* and the '*Ethics Policy Governing Research Involving Human Participants, Personal Data and Human Tissue'* (Ethics Policy) and, where externally funded, with the terms and conditions of the research funder.

In signing this research ethics application form I am also confirming that:

- The form is accurate to the best of my knowledge and belief.
- The project will abide by the University's Ethics Policy.
- There is no potential material interest that may, or may appear to, impair the independence and objectivity of researchers conducting this project.
- Subject to the research being approved, I undertake to adhere to the project protocol without unagreed deviation and to comply with any conditions set out in the letter from the University ethics reviewers notifying me of this.
- I undertake to inform the ethics reviewers of significant changes to the protocol (by contacting my academic department's Ethics Administrator in the first instance).
- I am aware of my responsibility to be up to date and comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data, including the need to register when necessary with the appropriate Data Protection Officer (within the University the Data Protection Officer is based in CiCS).
- I understand that the project, including research records and data, may be subject to inspection for audit purposes, if required in future.
- I understand that personal data about me as a researcher in this form will be held by those involved in the ethics review procedure (e.g. the Ethics Administrator and/or ethics reviewers) and that this will be managed according to Data Protection Act principles.
- If this is an application for a 'generic' project, all the individual projects that fit under the generic project are compatible with this application.
- **I understand that this project cannot be submitted for ethics approval in more than one department, and that if I wish to appeal against the decision made, this must be done through the original department.**

Name of the Principal Investigator (or the name of the Supervisor if this is a postgraduate researcher project):

Dr. Charith Abhayaratne

If this is a postgraduate researcher project, insert the student's name here:

Mr. Sandipan Pal

Signature of Principal Investigator (or the Supervisor):

	Date:

Email the completed application form and provide a signed, hard copy of 'Part B' to the Ethics Administrator (also enclose, if relevant, other documents).

The University Of Sheffield.

Consent Form for Participation in the focus group for the project Adaptive Lifestyle Monitoring

Please tick all the boxes.

1. I have read and understood the Information booklet about the study and I have had the opportunity to ask questions about the project. ☐
2. I understand that my participation is voluntary. If I wish to not participate in the study group, then I would notify the researcher at least 48 hours prior to the meeting date. During the focus group, should I not wish to answer any particular question or questions, I am free to decline. ☐
3. I agree that the proceedings of the focus groups would be audio-taped and transcribed. The transcription files would be securely kept in University archives. Only the researchers of the project (PhD student and the supervisors) would have access to the archives. ☐
4. I understand that the information I provide will be kept strictly among the researchers and the group that I participate in. The information will be securely stored and access to my responses will be restricted to the researchers working on this project. After the conclusion of the project, the transcription files would be kept for a period of 3 years within the University archives and only the researchers previously associated with the project would have access to them. ☐
5. I understand that my identity will not be revealed in any report produced out of the research. ☐
6. Once I have participated in the focus group, I agree to any information I provide being used for the purposes of the study and kept in the University archives for 3 years after the project concludes. ☐
7. I agree to take part in the study (focus group). ☐

Participant	**Researcher**	**Witness**
Signature	Signature	Signature
Name	Name	Name
Date	Date	Date

One signed copy to be kept by the participant and one copy by the researcher.

The University Of Sheffield.

FUTURE MONITORING TECHNOLOGY – A group discussion

Do you enjoy living independently and safely?

Are you interested in using technology?

Have you heard about video-based technologies?

Do you know that the camera is slowly becoming intelligent now-a-days?

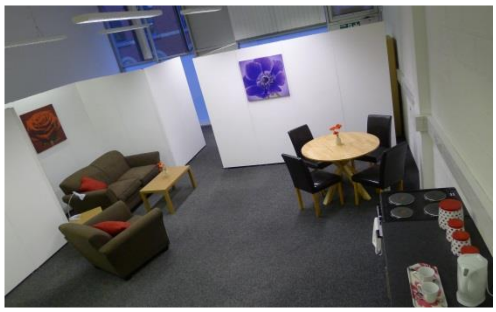
Come and participate in the focus-groups.

As a part of a research project within The University of Sheffield, titled '***Adaptive Lifestyle Monitoring***', we are welcoming people to participate in a focus group to discuss about issues regarding video-based monitoring technologies. It is a fantastic opportunity to voice out your concerns and thoughts about a technology of the future over a cup of tea and cakes.

The discussion would last about 2 hours and would be held within the working hours of a weekday.

Where will the discussion be held?

HOMELAB,
The University of Sheffield.
The Innovation Centre,
217, Portobello Street,
Sheffield, S1 4DP



Read through this Information Booklet carefully to know more about the research study

Contact Us: Mr. Sandipan Pal (PhD Student) : 0 794 643 5948, s.pal@sheffield.ac.uk
Dr. Charith Abhayaratne (Lecturer) : 0 114 222 5893, c.abhayaratne@sheffield.ac.uk
Prof. Mark Hawley (Professor) : 0 114 222 0682, mark.hawley@sheffield.ac.uk

The University Of Sheffield.

Information Booklet for the participants of the focus group

You are being invited to participate in a group discussion for a project named ***Adaptive Lifestyle Monitoring***. The main purpose of the study is to explore the relevance of a video-based monitoring system and also the privacy concerns of having a camera/visual sensor within the house. This study is a part of a PhD project in order to understand the context and relevance of a video-based lifestyle monitoring system.

Please read through this information booklet carefully before deciding whether to participate in the study.

What is the purpose of this research?

The alarming increase in the elderly population in the next 50 years envisages an increasing need of technology to support the daily living of elderly people. Extensive research is being carried out all around the world to develop these technologies for a better and secure future. These technologies are often termed as ‘Assisted Living technologies’ or ‘Lifestyle Monitoring technologies’. Lifestyle monitoring technologies are information and communication technology to assist, improve and monitor the daily living of the old and vulnerable population. The three main aims of such a technology are

- promote greater independence
- provide a safe and secure environment
- reduce the health-costs

Research has shown that there might be a link between the health of a person and the daily activities that is being undertaken. The technologies are deployed to monitor the activities and draw conclusions to foresee a potential decline in the health condition. Most of the present lifestyle monitoring technologies is typically sensor-based solutions, where a number of embedded or body-worn sensors are deployed or connected over a network to monitor activities of an individual. In this research we would want to explore the use of camera for monitoring the activities.

What is a camera-based solution?

A camera-based solution is one where a single or multiple cameras take images/videos of the room and the computer automatically analyses those images to deduce useful information for the carers or family. Camera-based solutions are becoming increasingly popular in various domains of our living like surveillance (CCTV footage), creative media industry, automated traffic management etc. A camera-based solution for a monitoring technology is a next generation technology. With your participation in this study, you would be able to contribute your views for a social technology of the future.

What are the advantages of using a camera instead of sensors?

It is said that ‘A picture is worth a thousand words.’ So having a picture or a video sequence of a particular scenario would tell us more about the situation.

A relatively fewer number of cameras are needed to installed within the dwelling instead of a complex network of sensors.

Visual data (images/videos) is contextually rich than sensor data. Recorded video sequences would also help the individual to learn more about their own daily living and what all changes would be needed for their own safety.

What is the purpose of this study?

Monitoring technologies often collect important information about ones’ daily living. Though the fundamental advantage of using a camera is the fact that a lot of information can be gathered from images and videos, however images and videos of ones’ daily living can be intrusive. This study is about understanding the usability and acceptance of a camera-based monitoring technology. Images/ Video sequences are extremely sensitive data and we would like to know about the privacy concerns of the users of such a technology.

Who can take part in the research?

Any individual above the age of 70 years can take part in this study. We would also be running 1 group discussion session for participants between the age of 45 and 70 years of age.

There is no restriction apart from the age bracket to participate in the study. However it must be noted the proceedings of the discussion would be audio-recorded and so anyone with severe speech impairment would find it difficult to participate in the discussion.

The participants would need to be physically mobile and relatively healthy in order to commute to the location of where the focus group would be held. Adequate English speaking skills are required in order to understand and follow the instructions.

All participants are asked to sign a consent form before participation. Your participation is voluntary and if you wish to withdraw, then you would need to notify any of the researchers at least 48 hours before the agreed meeting date.

No specific skill or technical expertise is needed to participate in the discussions.

How long would the discussion last?

The group discussion will be about 1.5 hours to 2 hours long. Attached is the plan of the different stages of the focus group. All the discussion proceedings will be audio-recorded.

What do I have to do?

You are being asked to participate in a focus group to discuss among a group of 6 to 8 participants about the proposed system and comment if such a system would be useful in the future for better and independent living. A focus group is a group discussion where the participants comment about a particular subject of research based on their personal experiences and understanding. Once you agree to participate in the discussion, you would need to sign a consent form.

On the meeting date, you will be shown short presentations about the system. You will also be shown some demonstration of some other camera-based applications to give you an idea of the usefulness of a video-based system. Following the presentations, there would a discussion among the participants regarding what you think of such a camera-based monitoring system and whether you think such a system would be helpful for independent living. From the discussion the researchers intend to understand the concerns and different perspectives of such a system. Your participation would be only for about 2 hours for 1 day.

Light refreshments (tea, coffee, biscuits) would be available during the discussions.

[Please note that you are not being asked to evaluate the system. The study is to understand the relevance of such a system.]

The University Of Sheffield.

Where the group discussion would be held?

We intend to hold the discussion in the brand new HOMELAB within the University. The HOMELAB is an amalgamation of a laboratory and a typical home or care environment. It is being setup by the new Centre (CATCH) of the University. To learn more about CATCH and the HOMELAB, please visit (http://www.catch.org.uk/).

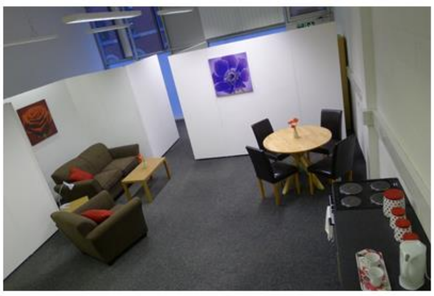
The address for the HOMELAB is –

The Innovation Centre,
217 Portobello Street,
Sheffield S1 4DP

If the HOMELAB is unavailable, then a meeting room within the University will be booked.

The date and the time for the discussion are negotiable based on the availability of the participants and the venue. The discussion would definitely be held during the normal working hours of a normal weekday.

[Please note that the venue for the group discussion would definitely have the *'ease of access'* option]



I am not confident to travel alone. Can my carer come along?

Yes. You can bring a carer along with you. However the carer will not be able to participate in the discussion. Also the researchers need to be pre-notified if you are bringing a carer along with you.

[Reimbursement of the travel cost is negotiable]

What is the agenda for the day?

Your participation would be for about 2 hours for 1 day. The detailed agenda is as follows-

Stage I : Introduction (15mins)

1. Welcome
2. A brief introductory presentation about the study
3. Individual introductions

Stage II : The opening discussion (20mins)

1. Presentation on Video-based monitoring Technology and its advantages (demos of gesture recognition or other computer vision based examples would be shown. The demos are intended to gain confidence among the participants about the recent advancements of camera-based technologies in different domains.)
2. Discussion

Stage III : Demo and discussion(40mins)

1. A brief demonstration of a video-based system (video/animation based)
2. Discussion
    a. What sort of activities/events should the technology identify for their profile?
    b. Where and how would they like the video information to be stored?
    c. To whom should this information be disclosed to? Would they like to control it?

Stage IV : Concluding remarks (10 mins)

[All timings are approximate]

How would the data be collected?

All your views would be audio-recorded. You would not have to identify yourself while speaking as all personal details would be confidential. The audio-recordings would be transcribed and analysis by the researchers. The audio-recording files and the transcription files would be stored within the University archives.

During the discussions, either the moderator or the orator might also take notes about the proceedings. All the notes and transcribed files would be used by the researchers for analysis.

What would happen to the audio recordings?

All audio recordings will be securely stored and used only for research purposes. The recordings will be transcribed and then analysed. Under no circumstances will the identity of the participants be revealed in any of the results or publication of the results.

After the conclusion of the project, the audio recordings and the transcribed files might be kept for a further period of 3 years within the University archives and only the researchers previously associated with the project would have access to them. 3 years after the conclusion of the project, the files would be deleted from the University servers.

Are there any risks or disadvantages to taking part in this study?

We do not anticipate that there will be any risks or disadvantages to taking part. You are being asked to participate in a group discussion and give your comments about the proposed technology.

Will my taking part in the study be kept confidential?

Group discussions are not entirely confidential as opposed to personal interviews. However all personal information will be kept strictly confidential among the researchers. Only the other participants in your group would know about your views but the audio-files or transcribed files will not be shared with them. Your personal details will be confidential only among the researchers and will not be shared with anybody. You will be identified by a code rather than a name. Audio recordings and the transcribed files will be stored securely within the University archives and used only for research purposes.

After the conclusion of the project, the transcription files would be kept for a period of 3 years within the University archives and only the researchers previously associated with the project would have access to them.

What will happen to the results of the research study and who is organizing and paying for the research?

The study is a part of a PhD project and the result will form part of the thesis of the researcher. The researcher would also aim to publish results of the study in a scientific journal and present our findings at professional conferences and events.

The PhD study is funded by the PIPIN network of The University of Sheffield. More details of the PIPIN (Promoting Independence for Personalized INteractive technologies) network can be found at http://pipin.group.shef.ac.uk/

I am ready to participate. What to do next?

If you are interested to participate in the discussion, then you would need to inform the researchers. Get in touch with the researchers over phone or email. The contact information are given below.

The University Of Sheffield.

Please notify the researchers whether there will be a carer accompanying you.

Before participation you would need to sign a consent form confirming your participation.

What are the potential benefits of participating in the study?

In recent years, camera-based technology has found its application in various domains like gaming, surveillance etc. Your participation in such a discussion would contribute to the research of future assisted living technologies.

What if something goes wrong?

If you wish to raise a complaint regarding your participation, then you should inform the Principal Investigator, Dr. Charith Abhayaratne.

Contact information -

For any other information, please contact -

Mr. Sandipan Pal
PhD Student
Department of Electronic and Electrical Engineering
The University of Sheffield
e-mail: s.pal@sheffield.ac.uk

You can also contact -

Dr. Charith Abhayaratne
Lecturer
Department of Electronic and Electrical Engineering
The University of Sheffield
e-mail: c.abhayaratne@sheffield.ac.uk
Phone: (+44) (0) 114 222 5893

Prof. Mark Hawley
Professor of Health Services Research
School of Health and Related Research
The University of Sheffield
e-mail: mark.hawley@sheffield.ac.uk
Phone: (+44) (0) 114 222 0682

Ethical Clearance The following document is the Ethical Clearance Certificate.

The University Of Sheffield.

Electronic & Electrical Engineering.

Mr Sandipan Pal
Department of Electronic &
Electrical Engineering
University of Sheffield

Dr Peter I Rockett, PhD
Sir Frederick Mappin Building
Mappin Street
SHEFFIELD
S1 3JD
United Kingdom
www.shef.ac.uk/eee

Telephone: +44 (0) 114 222 5589
Email: p.rockett@sheffield.ac.uk

PIR/HJL

19 June 2014

Dear Sandipan

Further to the Ethics Application you submitted and the revisions thereto, I am pleased to inform you that the reviewers are now content and the Ethics Approval is formerly granted.

Yours sincerely

Dr Peter I Rockett

The following slides were the slides shown to the focus group participants.

The University Of Sheffield.

PIPIN

FUTURE MONITORING TECHNOLOGIES – a group discussion

The University Of Sheffield.

PIPIN

WELCOME!!!

About me –

Sandipan Pal

~ Currently pursuing PhD

~ Takes interest in evolving technologies which have a social impact

~ Takes interest in sports and plays badminton regularly

About the Observer -

Tian Feng

~ Currently pursuing PhD on HDR image coding

Ruilong Chen

~ Currently pursuing PhD on video analysis for assisted living

The University Of Sheffield.

PIPIN

Agenda for today

Stage I – Introduction to the study

Stage II – Initial Discussion about monitoring technologies

Stage III – A prototype demonstration and discussion

Stage IV – Concluding remarks

Please help yourself with tea/coffee and biscuits

*We would aim to finish all the stages in 120 minutes

03/12/2017 3

The University Of Sheffield.

PIPIN

STAGE - I

✓ A brief introduction about the study

✓ Individual introductions

03/12/2017 4

The University Of Sheffield.

PIPIN

Introduction

Motivation –

To provide a long-term care facility to older people with a **camera-based monitoring system** thereby promoting greater independence among them.

Underpinning assumption –

There is a link between ones' health condition to the amount of activity undertaken by an individual

The University Of Sheffield.

PIPIN

Introduction

Today we will discuss about –

- Monitoring technologies
- Camera-based solutions
  - Safety
  - Privacy
  - Control of the technology

There are no right or wrong answers. The aim of this study is to find out how one feels about a video-based monitoring technology.

The University Of Sheffield.

PIPIN

Individual introductions

The University Of Sheffield.

PIPIN

STAGE - II

- ✓ Videos of intelligent video-based solutions
- ✓ Presentation on monitoring technologies
- ✓ Discussion

What is a camera-based solution?

- A video-camera records a scene or an action
- A computer automatically analyses the scene/action and gives an output

Where are camera-based solutions used?

- Traffic surveillance
- Crowd Management
- Computer games

What is a video-based solution?

- A video-camera records a scene or an action
- A computer automatically analyses the scene/action and gives an output

Where are camera-based solutions used?

- Traffic surveillance
- Crowd Management
- Computer games

Let us see an example where an intelligent camera automatically understands a scene and gives a relevant output

The University Of Sheffield.

What is a video-based solution?

- A video-camera records a scene or an action
- A computer automatically analyses the scene/action and gives an output

Where are camera-based solutions used?

- Traffic surveillance
- Crowd Management
- Computer games
- **Health Monitoring (still under research)**
  - Fall Detection

The University Of Sheffield.

Monitoring technologies and its evolution

What are monitoring technologies?

A technology which gathers information about ones’ daily living by collecting information actively or passively about the living conditions

Evolution of SMART homes –

The University Of Sheffield.

PIPIN

Camera-based monitoring technology

- A camera-based monitoring technology uses a camera and a computer to automatically understand about the living conditions of an individual.

- Why not sensor-based?
  - A lot of sensors are needed
  - Often complicated and obtrusive
  - Not enough information

03/12/2017 13

The University Of Sheffield.

PIPIN

Discussion

- How do you feel about such a technology?

03/12/2017 14

The University Of Sheffield.

PIPIN

Discussion

- How do you feel about such a technology?
- What do you think about camera-based solutions?

The University Of Sheffield.

PIPIN

Discussion

- How do you feel about such a technology?
- What do you think about camera-based solutions?
- What are your thoughts about a monitoring technology?

Discussion

- How do you feel about such a technology?
- What do you think about camera-based solutions?
- What are your thoughts about a monitoring technology?
- Do you foresee a video-based solution for a monitoring technology that you would like to use for yourself or a family member?

STAGE - III

✓ Demonstration of a prototype video-based monitoring technology

The University Of Sheffield.

PIPIN

STAGE - III

✓ **Discussion**

- ✓ What sort of activities/events should the technology identify?

The University Of Sheffield.

PIPIN

STAGE - III

✓ **Discussion**

- ✓ What sort of activities/events should the technology identify?
- ✓ Where and how would they like the information (video/health) be stored?

STAGE - III

- ✓ **Discussion**
  - ✓ What sort of activities/events should the technology identify?
  - ✓ Where and how would they like the information (video/health) be stored?
  - ✓ To whom should the information be disclosed to? With whom should the control to access the information be with?

STAGE - IV

- ✓ Concluding remarks

The University Of Sheffield.

PIPIN

STAGE - IV

✓ Concluding remarks

THANK YOU FOR YOUR TIME AND PARTICIPATION

03/12/2017 23