A Probabilistic Model for Assistive Robotics Devices to Support Activities of Daily Living

Mitesh Patel

Submitted in fulfilment of the requirement for the degree of Doctor of Philosophy

The University of Technology, Sydney

The Faculty of Engineering and Information Technology

Mechatronics and Intelligent Systems Group

Centre for Autonomous Systems

www.uts.edu.au

Supervisor	:	Dr. Jaime Valls Miró
Co-Supervisor	:	Prof. Gamini Dissanayake

June 2014

Declaration of Authorship

I, Mitesh Patel, declare that this thesis entitled '*A Probabilistic Model for Assistive Robotics Devices to Support Activities of Daily Living*' and the work presented in it, is my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work done by others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what contributions I have made to my thesis.

Signed:

Production Note:
Signature removed prior to publication.

Date: June 13, 2014

Acknowledgements

First and foremost, I would like to thank my thesis advisors, Dr. Jaime Valls Miró and Prof. Gamini Dissanayake, for never denying an opportunity to learn and grow as a researcher. Pursuing a Ph.D. wouldn't have been possible without their unending advice and support in all areas call it writing papers, doing research presentation, participating in different forums on world stage. They've been mentors and have always encouraged cross collaboration with other universities which has proven to be an invaluable experience during my time as a Ph.D. student.

I would like to thank my parents. They have always believed in the value of education. They let me choose my path, which eventually led me to enrol as graduate student at the University of Technology, Sydney (UTS).

I would also like to thank my wife for her encouragement, help and unflagging support. The many, many hours spent reviewing my slides, reading my papers, and "volunteering" for all the experiments is greatly appreciated.

I would also like to thank all the bright minds who were around me at the Centre of Autonomous Systems (CAS), especially Marc Carmichael, Michael Behrens, Alen Alempijevic, Andrew To, Gavin Paul and Tarek Taha for all the intelligent conversations, coffee breaks and "volunteering" for experiments, the results of which are collected in different chapters.

Contents

List of Tables

List of Figures

Abbreviations

ADL	Activities of Daily Living
AP	Action Primitives
DOF	Degrees Of Freedom
DBN	Dynamic Bayesian Network
EM	Expectation Maximisation
HRI	Human-Robot Interaction
HMM	Hidden Markov Model
HHMM	Hierarchical Hidden Markov Model
HDBN	Hierarchical Dynamic Bayesian Network
IAD	Intelligent Assist Device
L-DBN	Layered Dynamic Bayesian Network
POMDP	Partially Observable Markov Decision Process
SL	Structure Learning
UN	United Nations
WHO	World Health Organisation

Glossary

Activities of Daily Living (ADLs)	Activities of Daily Living are defined as a set of basic activities necessary for normal self-care and independent living
Human Robot Interaction (HRI)	Human Robot Interaction is a branch of robotics science that focuses on modelling, implementing and evaluating the collaboration between robotic systems and human partners
Learning by Imitation	Learning by imitation is an approach that has been used by roboticists for bootstrapping learning of robot activities based on human observation
Action Primitives (APs)	Action Primitives are pool of semantic structure that is generated by decomposing complex activity space into atomic actions

Abstract

A Probabilistic Model for Assistive Robotics Devices to Support Activities of Daily Living

This thesis explores probabilistic techniques to model interactions between humans and robotic devices. The work is motivated by the rapid increase in the ageing population and the role that assistive robotic devices can play in maintaining independence and quality of life as assistants and/or companions for these communities. While there are substantial social and ethical implications in this pursuit, it is advocated that robotic systems are bound to acquire more sophisticated assistive capabilities if they are to operate in unstructured, dynamic, human-centred environments, responsive to the needs of their human operators. Such cognitive assistive systems postulate advances along the complete processing pipeline, from sensing, to anticipating user actions and environmental changes, and to delivering natural supportive actuation. Within the boundaries of the human-robot interaction context, it can be expected that acute awareness of human intentions plays a key role in delivering practical assistive actions. This work is thereby focused on the human behaviours likely to result from merging sensed human-robot interactions and the learning gained from past experiences, proposing a framework that facilitates the path towards integrating tightly knit human-robot interaction models.

Human behaviour is complex in nature and interactions with the environment and other objects occur in different and unpredictable ways. Moreover, observed sensory data is often incomplete and noisy. Inferring human intention is thus a challenging problem. This work defends the thesis that in many real-world scenarios these complex behaviours can be naturally simplified by decomposing them into smaller activities, so *that their temporal dependencies can be learned more efficiently with the aid of probabilistic hierarchical models. To that end, a strategy is devised in the first part of the thesis to efficiently represent human Activities of Daily Living, or ADLs, by decomposing them into a flexible semantic structure of "Action Primitives" (APs), atomic actions which are proven able to encapsulate complex activities when combined within a temporal probabilistic framework at multiple levels of abstraction. A Hierarchical Hidden Markov Model (HHMM) is proposed as a powerful tool capable of modelling and learning these complex and uncertain human behaviours using knowledge gained from past interactions.*

The ADLs performed by humans consist of a variety of complex locomotion-related tasks, as well as activities that involve grasping and manipulation of objects used in everyday life. Two widely used devices that provide assistance to users with mobility impairments while carrying out their ADLs, a power walker and a robotic wheelchair, are instrumented and used to model patterns of navigational activities (i.e. visiting location of interest), as well as some additional platform-specific support activities (e.g. standing up using the support of assistive walker). Human indications while performing these activities are captured using low-level sensing fitted on the mobility devices (e.g. strain gauges, laser range finders). Grasping and manipulations related ADLs are modelled using data captured from a stream of video images, where data comprises of hand-object interactions and their motion in 3D space.

The inference accuracy of the proposed framework in predicting APs and recognising long term user intentions is compared with traditional discriminative models (sequential Support Vector Machines (SVM)), other generative models (layered Dynamic Bayesian Networks (DBN)), and combinations thereof, to provide a complete picture that highlights the benefits of the proposed approach. Results from real data collected from a set of trials conducted by actor users demonstrate that all techniques are able to predict APs with good accuracies, yet successful inference of long term tasks is substantially reduced in the case of the layered DBN and SVM models. These findings validate the thesis' proposal that the combination of decomposing tasks at multiple levels and exploiting their inherent temporal nature plays a critical role in predicting complex interactive tasks.

Chapter 1

Introduction

1.1 Research Problem

Robots have proven to be a powerful tool in predictable environments such as factories and manufacturing plants to improve productivity and perform dangerous or monotonous tasks. Lately, research has been focused on the potential of using robots to aid humans outside the strict 'industrial' environments such as hospitals, offices or home settings. However, technology is still, a long way away from producing robots capable of working alongside humans and demonstrating the same competencies as humans. This is due to the inherently complicated nature of human behaviour, and the limitations of sensors in capturing this complexity. It therefore remains an open challenge to model complicated human behaviour, such that the robotic system could better understand this behaviour and act in the role of an assistant or partner. *Human-Robot Interaction* (HRI) is a branch of robotics science that focuses on modelling, implementing and evaluating the collaboration between robotic systems and human partners. The increasing number of application domains in which robots can and will be deployed in the future, and the inevitable need to interact with humans in many of these domains are the motivating forces driving further developments in HRI.

Under the HRI umbrella, the robotic systems developed are such that the robot and human share a common environment and work in a symbiotic relationship to achieve a common goal. The interaction between human and robot should be both natural and effective so that the robot is able to detect the behaviour of the user on the basis of his/her motion, and is able to assist the human in achieving his/her goals. Considerable focus has been given to those robotic systems that includes humans *'in-the-loop'*. Such system works on the principle of a 'perceive-sense-act' mechanism to provide a better natural collaboration between robots and humans compared to the commercially available robots to-date. Researchers have been working in this area to develop a HRI system that accepts user commands in a natural way and assists users according to their needs which are socially acceptable thus allowing a robot to be a reliable personal companion and assistant.

1.2 Motivation

Demographic projections show that the world's ageing population is rapidly increasing. There is a marked demographic shift at a global level indicating that the worldwide proportion of people aged over 60 is expected to double between 2000 and 2050 [United Nations, 2006] (trends pictorially depicted in Fig. 1.1 by the Congressional Budget Office [Congressional Budget Office, 2005]). A recent report by the Australian Academy of Technological Sciences and Engineering further canvassed various options based on the use of emerging innovative technologies to address these challenges [Tegart, 2010]. The principle of the *'Convention on the Rights of Persons with Disabilities (CRPD)'* promoted by the United Nations states that a support service provided to a disabled person should be such that it enables keeping them within the community, and not in a segregated setting [Officer and Posarac, 2011]. Such support systems would give otherwise immobile or dependent people the freedom of movement which would significantly increase their independence and potentially improve their overall quality of life. The challenges associated with such support services, along with other health and longevity paradigms, are driving the need for improvements in a range of services related to aged care. Systems such as smart blind sticks [Kang et al., 2001], robotic wheelchairs [Carlson and Demiris, 2010; Demeester et al., 2006; Hoey et al., 2007; Mandel et al., 2005; Taha et al., 2008] smart robotic walkers [Alwan et al., 2005; Dubowsky et al., 2000; Hirata et al., 2006; Morris et al., 2003; Omar et al., 2010; Wasson et al., 2008] and robotics grasp manipulator [Saxena et al., 2008; Srinivasa et al., 2008] have been the subjects of significant research in the quest to achieve this goal.

The work presented in this thesis, focuses on modelling the Activities of Daily Living (ADLs) of elderly and disabled people, which is the first step towards the development of a tightly knit HRI system using assistive devices. According to the World Health Organisation (WHO), ADLs are defined as a set of basic activities necessary for normal self-care and independent living [World Health Organisation, 2004]. These activities comprises of 'movement in bed', 'transfers from sitting to standing position',' to and from toilet and bed', 'locomotion', 'dressing', 'personal hygiene' and 'feeding' [The Repatriation Commission, Australia, 1998; Veterans Affairs Canada, 2006]. ADLs are mainly used as assessment criteria to measure the level of disability/impairment present in a person, which is then used by physical and/or occupational therapists to prescribe an assistive device so as to compensate for the impairment. The prescribed assistive devices can ultimately provide the necessary support required to perform ADLs independently and to improve the overall quality of life. This emerging area of developing intelligent assistive support devices is more generically referred to as 'assistive robotics' which advocates devices having the ability to work collaboratively with their human users in the pursuit of the user's objective. According to the WHO, assistive technology is defined as 'an umbrella term for any device or system that allows individuals to perform activities they would otherwise be unable to do, or increase the ease and safety with which activities can be performed' [World Health Organisation, 2004]. The term

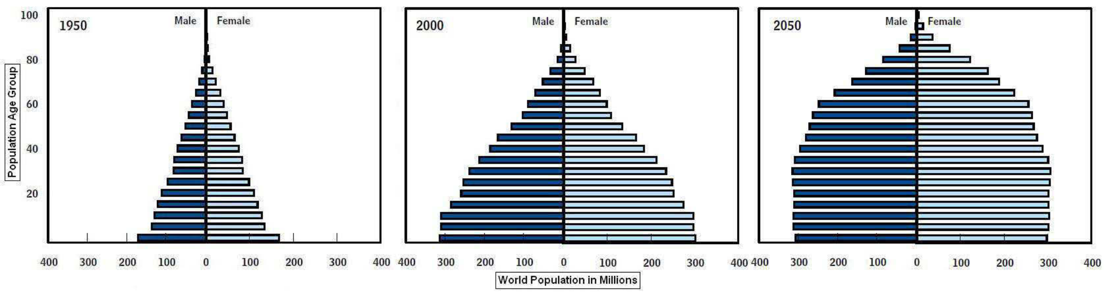


Figure 1.1: Changes in the Population Structure [Congressional Budget Office, 2005]

'assistive technology' goes beyond the simplistic scenario of providing automatic services at the press of a button or without taking account of the human in-the-loop; rather, it captures the concept of robotic machines being able to assist people in achieving their goals of performing ADLs by intelligently sharing control between them.

This concept of a shared control strategy is not new, and is evident in the design and application of many technological tools in use today. For instance, the *da Vinci*© Surgical robotic system [1] allows doctors to perform surgery with the assistance of a robot, and has been used in American operating rooms since 2000. In the car industry, adaptive cruise control in high-end Daimler-Benz and BMW cars uses a forward-looking radar to detect the speed and distance of the vehicle ahead, automatically adjusting in response to the user-controlled speed in order to maintain a proper distance. Ultrasonics and cameras which are employed as parking aids, as well as blind-spot detectors work in collaboration with the driver to increase safety and comfort. These examples of assistive technologies are all readily available today. Yet despite the enormous potential of this field of research to address the challenges that are emerging due to rapid growth of the ageing population, their application to the domain of the elderly and disabled is severely limited.

1.3 Approach and Methodology

In the scope of this research work, some assumptions have been made so as to facilitate the research question in an more efficient manner. The individual is assumed to need assistance to perform some of the many daily activities. We also assume that despite having partial disability, the human is still able to manoeuvre around with some external support from a person or an assistive device like a walker, wheelchair or cane, and is able to perform some of the many ADLs with minimal support. The individual is assumed to behave as naturally as her or his capabilities allow when attempting to perform an activity. As the core of this research is understanding human behaviour from the perspective of natural interaction, we assume that there is no explicit interface

[1] www.intuitivesurgical.com

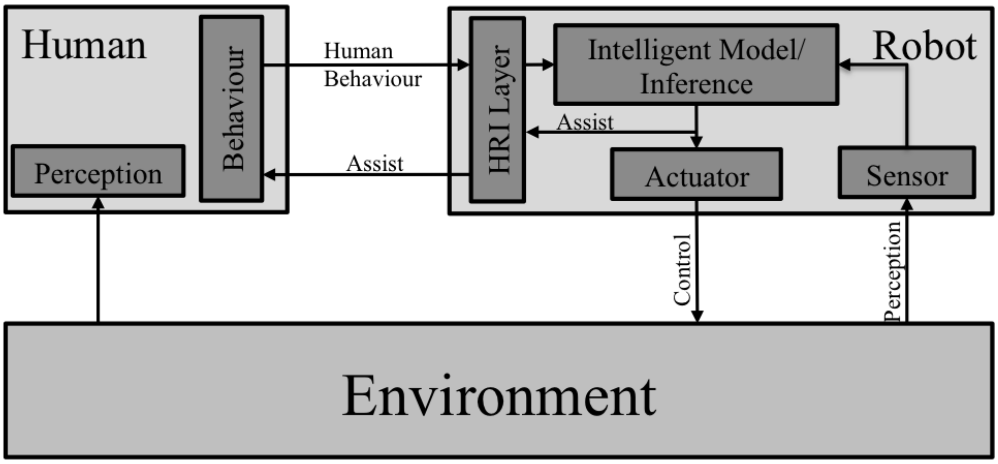


Figure 1.2: Interaction between a human, a robotic agent and the environment

like speech or a switch button between the human and the robot, as this would be an unnatural way of communicating. The user's behaviours are perceived by the robot via different sensor interfaces present on the robot.

The ADLs performed by the target group used in our research work are both complex and diverse. The ADLs consist of a pool of activities which can be both short-term in nature (*e.g.support to stand up*) or long-term (*e.g. take me to the bedroom, assist me in pouring water into a mug*). Modelling ADLs by simplifying the natural complexity present in ADLs makes the process of perceiving and understanding human behaviour more simplistic for the robot. The combination of modelling and learning ADLs proposed in the methodology in this thesis is aimed at providing a natural human-robot collaborative mechanism. Figure 1.2 depicts the interactions between the human, robot, environment and the intelligent system, which is the subject of the research described in this thesis.

Firstly we deal with the research question of representing ADLs performed by the elderly in their everyday life. The ADLs performed by the elderly population are complex in nature and consists of activities such as locomotion, dressing, feeding, personal hygiene [The Repatriation Commission, Australia, 1998; Veterans Affairs Canada, 2006]. These ADLs can be performed at different locations and at different time of the day. Complexity is further increased by the individual differences in the behaviours of people performing the same activity. These complexities can become intractable with the addition of more and more activities. This issue motivated us to develop a flexible semantic structure called 'Action Primitives (APs)' that can represent the complex action space by decomposing it into atomic actions. Representation of ADLs using a dictionary of sub-actions becomes an efficient solution as it minimises the action space into smaller clusters of APs. This concept has been utilised in a number of application such as grasping and manipulation of objects [Krüger et al., 2010], learning activities from human demonstration [Lee et al., 2013], planning navigational task [Liao et al., 2003] etc. Using a dictionary of APs to define ADLs, gives the advantage of scalability and re-usability as the entire problem space can be reduced to a number of APs which can be reused in different sequence to define any given ADL.

The second part of this research work focuses on the problem of modelling ADLs and the associated APs. We deploy a probabilistic framework capable of modelling the uncertainty and complexity present in human behaviour during the course of everyday activity.

Finally, we evaluate our proposed approach to model and infer some of the many basic ADLs which are necessary for independent living of a person. We model locomotion related ADLs, which involves visiting the location of interest, support activities such as transfer from standing to sitting position or vice-versa and activities related to the manipulation of everyday objects. To model locomotion related ADLs we collected user data; while the user performs ADLs using the support of a power walker and a robotic wheelchair (shown in Figure 1.3) (details of both the platforms given in Appendix A and B). The data consist of human behaviour as perceived by different physical sensors fitted on the mobility devices. The sensors record different behaviours such as the readiness of the user to perform an ambulation activity or any other support activity (e.g. standing up). For modelling ADLs related to the manipulation of everyday objects, the data consist of hand-object interaction and its motion tracking captured through a stream of image frames by a RGB-D kinect sensor.

(a) Walker (b) Wheelchair

Figure 1.3: Instrumented rollator walker and wheelchair used in our experiment.

1.4 Contributions

The main contributions arising from this thesis work are summarised below:

- Propose a novel approach to bridge the gap between low level sensor measurement and high level everyday activities. ADLs performed by the elderly involve visiting different locations of interest in a home environment, standing up/sitting down using the support of the assistive device, or performing grasping and manipulation related activities. To achieve this we develop an action grammar based language which segregates complex ADLs into a string of APs which are sequenced in different combinations to define the ADLs.

- Outline a Hierarchical Hidden Markov Model (HHMM) as a unified probabilistic framework, which is capable of modelling APs (primitive human behaviours) and mapping them to the subsequent high level ADL performed by the user. The hierarchical nature of the framework enables an efficient representation of an ADL and its corresponding APs, which are inferred at a number of intermediate layers.

- Evaluate the proposed approach using two widely used mobility aid devices, a power walker and a robotic wheelchair as depicted in Figure 1.3. Additionally, we show that using the proposed methodology makes the system capable of inferring a variety of ADLs which can be static/support activities and/or navigational activities. The proposed approach of deploying a HHMM based framework to model ADLs at different levels of hierarchy out performs against other temporal models which does not possess hierarchical characteristics, thereby reinforcing that the hierarchical nature of the model makes it capable of exploiting the strong relationship between APs and ADLs.

- Extend the applicability of our approach to model ADLs related to grasping and manipulation of everyday objects. The evaluation of the probabilistic model on a different problem domain proves the versatility of our approach, which address the challenge of predicting a wide range of basic ADLs from low level sensor measurement.

1.5 Publications

The publications resulting from the work presented in this thesis are:

- "A Probabilistic Approach to Learn Activities of Daily Living of an Intelligent Mobility Aid Device User", Miró J.V., **Patel, M.**, and Dissanayake, G., *Robotica (Special Issue on Rehabilitation Robotics and Human-Robot Interaction)*, (Submitted)

- "A Probabilistic Approach to Learn Activities of Daily Living of a Mobility Aid Device User", **Patel, M.**, Miró J.V., and Dissanayake, G., *IEEE International Conference on Robotics and Automation (ICRA 2014)*, (to Appear)

- "Learning Object, Grasping and Manipulations Activities using Hierarchical HMMs", **Patel, M.**, Miró J.V., Kragic, D., EK, C. H.,and Dissanayake, G., *Journal of Autonomous Robots (Special Issue on Beyond Grasping: Modern Approaches for Dexterous Manipulation)*, (to Appear)

- "Language for Learning Complex Human-Object Interactions", **Patel, M.**, EK, C. H., Kyriazis, N., Argyros, A., Miró J.V. and Kragic, D., *Proc. of the IEEE*

International Conference on Robotics and Automation (ICRA 2013), pp. 4982 – 4987, 2013.

- “Probabilistic Activity Models to Support Activities of Daily Living for Wheelchair users”, **Patel, M.**, Miró J.V., and Dissanayake, G., *Proc. of workshop on Progress, Challenges and Future Perspectives in Navigation and Manipulation Assistance for Robotic Wheelchairs, IEEE International Conference on Intelligent Robots and Systems (IROS 2012)*, 6 page, 2012.

- “A Hierarchical Hidden Markov Model for Inferring Activities of Daily Living with an Assistive Robotic Walker”, **Patel, M.**, Miró J.V., and Dissanayake, G., *Proc. of the 4th IEEE RAS/EMBS International Conference on Biomedical and Biomechatronics (Biorob 2012)*, pp. 1071 – 1076, 2012.

- “Activity Recognition from the Interactions Between an Assistive Robotic Walker and Human Users”, **Patel, M.**, Miró J.V., and Dissanayake, G., *Proc. of The 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2011)*, pp. 221 – 222, 2011.

- “Probabilistic Models versus Discriminate Classifiers for Human Activity Recognition with an Instrumented Mobility-Assistance Aid”, **Patel, M.**, Khushsuaba R., Miró J.V. and Dissanayake, G., *Proc. of The Australasian Conference on Robotics and Automation (ACRA 2010)*, 8 page, 2010.

- “Dynamic Bayesian Networks for Learning Interactions between Assistive Robotic Walker and Human Users”, **Patel, M.**, Miró J.V., and Dissanayake, G., *Proc. of the 33rd Annual German Conference on Artificial Intelligence (KI 2010)*, pp. 333 – 340, 2010.

- “Stochastic Models for interactive human-robot assistive agents”, **Patel, M.**, Miró J.V., and Dissanayake, G., *Proc. of the Young Pioneers Workshop, 5th ACM/IEEE International Conference on Human-Robot Interaction(HRI 2010)*, 2 page, 2010.

- “Robotic Assistance with Attitude: A Mobility Agent for Motor function Rehabilitation and Ambulation Support", Miró J.V., Osswald V.,**Patel, M.**, and Dissanayake, G., *Proc. of the 11th IEEE International Conference on Rehabilitation Robotics (ICORR 2009)*, 529 – 534, 2009.

1.6 Thesis Overview

- **Chapter 2**

  In this chapter we outline the importance of ADLs and provide a brief background on work done by researchers in the area of activity recognition in general. We describe our proposed approach of representing ADLs using a dictionary of APs and the advantages associated with this approach. Further we revise the Bayesian based probabilistic framework, in particular the Dynamic Bayesian Network (DBN) and its Hierarchical variants, in order to provide the reader with a background on the graphical models used, and to introduce the notations.

- **Chapter 3**

  In this chapter, we focus on the everyday activities performed by a typical walker user. ADLs consists of support activities (*e.g. stand up*) and navigational activities involving manoeuvring to different locations of interest in a given indoor environment. We outline the details of various activities performed by a walker user and collect data while users' perform these ADLs. We further provide details of action primitive semantics, which is developed by dividing the user environment into a structural topological map. The topological map representation of the environment consists of junction points and edges which connect different locations of interest. These junction points act as APs which provide the necessary navigational cues to describe the overall ADL. We conclude this chapter with a description of the DBN and HHMM frameworks used to model ADLs and the corresponding APs. We evaluate and compare the inference accuracy of the HHMM model with a layered DBN model and a HHMM/SVM hybrid model.

- **Chapter 4**

  In this chapter, we shift our focus towards utilising intrinsic human motion to model ADLs. Motion related human behaviour is defined as the most atomic activity (based on the smallest meaningful change in user motion) in which ADLs can be decomposed. The atomic activities based on human motion which act as APs are then mapped to different navigational and support cues to model the overall ADLs. The primary advantage of using this approach is to avoid the dependency on topological maps to model ADLs. We demonstrate how the use of topological maps can be avoided by modelling ADLs using human motion models. We use the data collected while different users perform ADLs using two mobility devices: a power walker and robotic wheelchair. The dictionary of APs is based on the intrinsic human behaviours that occur when the use performs different ADLs. We detail a list of APs which can be combined in different sequence to describe the ADL a user is trying to perform. Finally, we describe the HHMM framework used to model APs and the associated ADLs, and conclude this chapter by listing the results of the HHMM framework and compare them with that of a Layered Dynamic Bayesian Network (L-DBN) and a two stage Support Vector Machine (SVM) classifier.

- **Chapter 5**

  In this chapter, we focus on understanding ADLs related to grasping and manipulation of objects used in everyday life. We apply a similar technique to that used in Chapter 3 and 4 whereby we explore the action grammar present underneath these ADLs. We extend the usage of the HHMM framework to model grasping and manipulation related ADLs by decomposing then into strings of meaningful APs. The data features used for our experiments consist of hand and object motion in the 3D space extracted from videos recorded using RGB-D kinect sensor. Along with hand-object tracking features, the data also consists of the features of each finger joints (such as the yaw and tilt angles of each joint). We conclude this chapter by evaluating the inference accuracy of the HHMM framework and comparing its performance to that of a HHMM/SVM hybrid model.

- **Chapter 6**

  The final chapter summarises the findings of this research and presents the conclusion that are drawn from this work. Future research directions are also outlined such as learning human behaviour and adapting to the gradual change in their behaviour, learning the structural dependencies of the model from data.

Chapter 2

Background

2.1 Introduction

Our work in this thesis draws on the rich foundation of work in robotics, machine learning, cognitive science and artificial intelligence. With the constant improvement in technology, robots are expected to be used in shared spaces with humans in the future. Projects have been presented which study different applications such as a robotic tour guide in a museum [Thrun et al., 1999], a robotic system for assisting people with dementia during hand washing [Hoey et al., 2007], a robot approaching potential customers in front of a shop [Kanda et al., 2009] and a robot used to provide physical guidance and support for navigating around in a environment [Dubowsky et al., 2000]. The problem tackled in most of these projects focused on technical issues like localisation and path planning or giving extrinsic instruction to the user.

In contrast, a major component involved in the design of a human-centered robotic system, where a robot acts as an 'assistant or helper' is the ability of the robots to intrinsically understand human behaviour. More research is needed into developing methods that accept user commands in a natural way. With the demographic shift in the population structure there has been considerable interest in developing and utilizing robots in elderly care. The overall aim of such robotic systems is to give a sense of independence which eventually leads to improvements in the overall quality of life for people with disability [Brose et al., 2010]. Researchers have focused on developing assistive robots to cater for different areas where assistance might be required to perform Activities of Daily Living (ADLs).

In this chapter, we focus on the importance of ADLs and the challenges associated with modelling these activities. We provide details of our proposed architecture and its applicability to modelling ADLs. This chapter also reviews the probabilistic approach which has become more popular in tackling the problem of activity recognition. We detail the Bayesian Network (BN) framework, in particular the Dynamic Bayesian Network (DBN) and the Hierarchical Hidden Markov Model (HHMM). This is reviewed in order to provide the reader with a background of Bayesian networks and the mathematical notations and formulations used in this thesis.

2.2 Activities of Daily Living

The term Activities of Daily Living is defined as a set of activities necessary for normal self-care. The activities comprises movements in bed, transfers, locomotion, dressing, personal hygiene, and feeding [The Repatriation Commission, Australia, 1998] [World Health Organisation, 2004]. It is mainly used as an assessment measure which helps practitioners/occupational therapist to determine how independent patients are and what skills they posses to accomplish activities on their own. The goal of this evaluation is to determine how well the patient can perform each of these activities and what sort of support/assistance (if any) would be required so as to function as independently as possible.

Predicting user activities/behaviour and assisting them in performing those activities is not a new concept and has been addressed in the context of many applications. Typical examples that utilise intention recognition to enhance the application at hand can be found in motion prediction [Luber et al., 2010], video surveillance [Nguyen et al., 2005], behaviour recognition [Kluge et al., 2001], gaze tracking [K. and Ramakrishna, 1999] and activity prediction [Demeester et al., 2006] [Buettner et al., 2009].

The central idea of developing an activity recognition system is to increase the level of adaptation in intelligent systems. In smart home intelligent systems, the monitoring is done in a passive manner. Sensors such as Radio Frequency Identification (RFID) [Buettner et al., 2009], a camera based tracker [Nguyen et al., 2005] [Kawanaka et al., 2005] installed in a given indoor environment are used to track human activities and train models that can later be used to predict human activities such as entering a room, using a telephone, brushing teeth etc. Buettner *et. al.* describe a dense sensing approach that uses RFID sensor network technology to recognise human activities [Buettner et al., 2009]. In the context of outdoor environments, Liao extracted the activities of a person and used this activity knowledge to predict places the person has visited from GPS data logs [Liao et al., 2007]. Nguyen *et. al.* used a probabilistic framework to model and recognise human activities in a confined space, while tracking the user with two static cameras [Nguyen et al., 2005].

2.3 Challenges of Recognising Activities of Daily Living

The importance and significance of recognising ADLs has been recognised for decades by the robotics, vision and computer science communities. In order to achieve the goal of developing a tightly knit Human-Robot interactive system that can assist users in performing everyday activities, it is important that the system can exploit meaningful information from user behaviour. Further, daily activities encompass a wide range of activities including for instance visiting different locations of interest or performing activities which involve grasping and manipulation objects.

In order to address the challenge of modelling high level ADLs from low level sensors, we develop a dictionary of human behaviours called Action Primitives (APs) which is used as an intermediate step to infer the high level activities. The concept of APs is not new and was inspired by the research done on human motion and other biological movements which postulates that movement behaviour is composed of simple, atomic movements that can be combined and sequenced to form complex behaviour [Ijspeert et al., 2002] [Schaal et al., 2003] [Kulic et al., 2011] [Lee et al., 2013]. APs allow

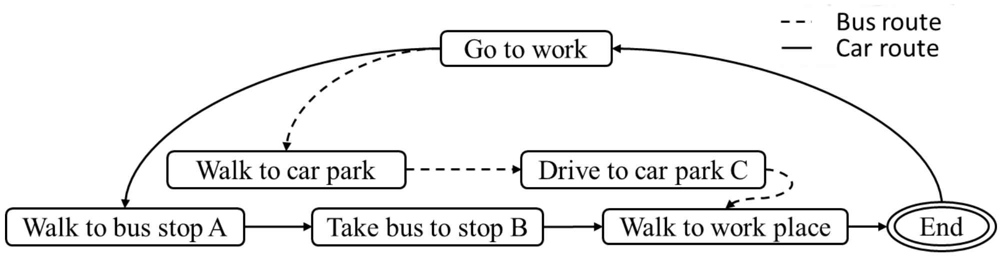


Figure 2.1: Activity of going to work which has two options, a bus route and a car route. Solid lines indicates decomposition of the activity into APs for taking the bus route and the dotted lines indicates the same for taking the car route

for complex activities to be decomposed into alphabets of primitives, which in-turn facilitates for a construction of grammar-like description to govern their order of use. The grammar-like structure underneath ADLs provides the motivation to develop a language of APs which represents the smallest meaningful units of action components. Once defined these APs can then be combined in different order to describe any complex ADLs. This approach provides a better understanding of more complicated ADLs, the structure of which can be generated by reusing APs in different sequences.

Further, as per the literature of psychophysiology, APs are present at several different hierarchical levels in human brain [Rizzolatti et al., 2001]. We utilise human brain like structure to represent APs at different levels of hierarchy to model higher-level ADLs.

2.3.1 Representing ADLs using Dictionary of APs

Almost every activity (if not all) performed by humans, consists of a structure underneath which can be exploited to predict the overall ADL. The pool of APs generated by decomposing and ADL is such that it represents the most atomic activities into which a given ADL can be decomposed. The number of APs in a dictionary will differ based on the type of ADL and the complexity of motion or movement involved in performing this ADL.

To give a few examples, in an outdoor navigation problem, the task of going to the office can be decomposed into string of APs shown in Figure 2.1. The string of APs can be used to represent two independent navigational routes: one followed when using a car and other followed when using a bus (shown in Figure 2.1). The activity of pouring water from a bottle can be decomposed into APs shown in Figure 2.2. To construct a guiding navigational route to a user, the Global Positioning System (GPS) subdivides the entire route into directional navigational cues to reach the end destination (example depicted in Figure 2.3). Similarly, in an indoor environment, activities performed using the support of a walking frame can broadly be decomposed into string of primitive actions as shown in Figure 2.4.

It should be noted that the ADLs in all the given examples above occur over a time period and hence consist of two types of information, spatial information and temporal information [Raptis et al., 2008]. Spatial information corresponds to the information that is obtained at a given instant of time, where as temporal information is related to the information collected over a time period. In our work, we define APs as a cluster of information which occurs over a time period and hence includes both spatial and temporal information. Each cluster of APs encodes a single dimension of a commonly occurring deformation. To explain this better, consider the example of pouring water from a mug (Figure 2.2) which is decomposed into a string APs. In this example the AP of *Approach* consists of the entire trajectory which starts from the hand at rest until it reaches the grasping position. Similarly, in case of AP of *Stand Up* (Figure 2.4), the AP consists of the entire motion which starts from the sitting position until the user is in a fully standing position.

Apart from the representation of high level activity using a sequence of APs, the representation is also advantageous in terms of *scalability*, whereby a complex activity can be represented using a set of defined APs, and *re-usability*, where APs are re-used in different sequences to construct any given activity. For example, in the representation

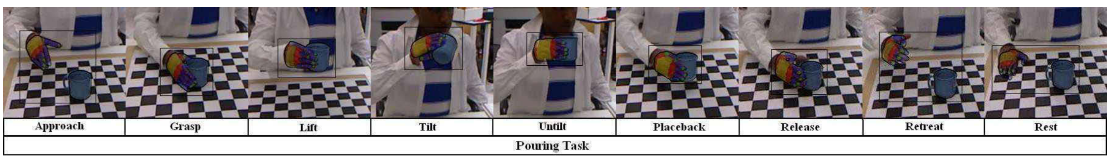


Figure 2.2: Activity of pouring water from a mug decomposed into a sequence of action primitives

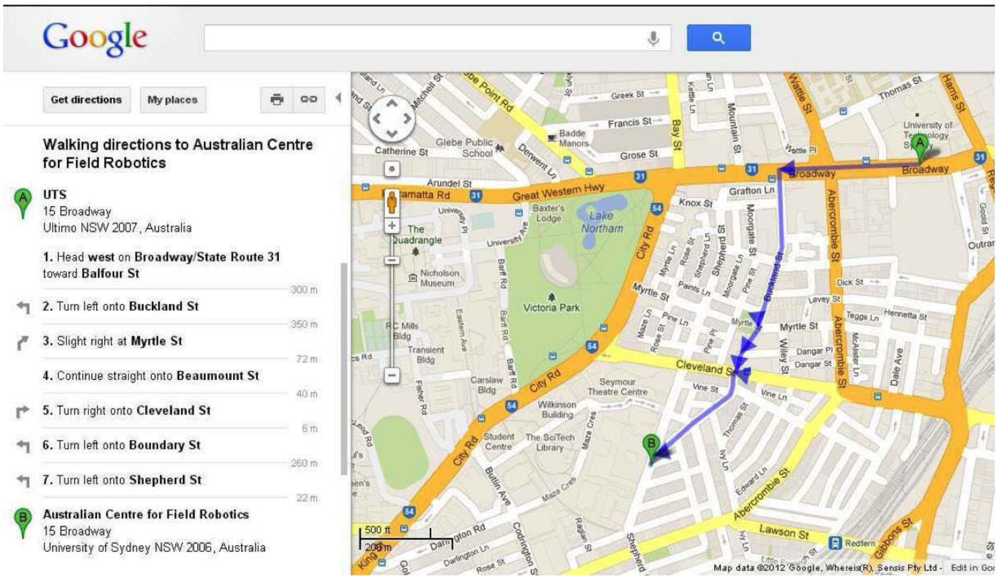


Figure 2.3: An example of a navigational activity decomposed by a GPS navigational system into directional navigational cues to guide a user to reach his/her intended destination

of the activity of going to office (Figure 2.1) the APs of 'walk to workplace' or 'walk to bus stop A' can be reused to define other activities. In the example of visiting a location of interest in an indoor environment, the APs of 'stand up' or 'go straight south' can be reused to define various other activities which might involve visiting other location of interest. The proposed architecture of the $ADLs - APs$ representation proves to be an efficient structure as it significantly reduces the state search space used to define ADLs.

Despite the efficient representation of ADLs using APs, other parameters such as the inherent uncertainty present in human behaviour and noisy sensor data add more ambiguity to the overall ADLs. This ambiguity makes it difficult if not impossible to model ADLs in a deterministic manner. In this respect, we choose probabilistic models, as our representative framework due to (1) their robustness of accommodating human ambiguities as a result of their probabilistic nature, (2) their ability to capture both spatial and temporal variability present in the APs and ADLs, in particular to capture the change in variance along the observation to model APs and variance in APs to model the overall ADLs, and (3) compactness on representing hierarchical and recursive structures.

Figure 2.4: Activity of visiting location of interest in an indoor environment using the support of a walker

2.4 Probabilistic Models

The use of probabilistic models has been widely investigated [Carberry, 2001] [Pynadath, 1999] [Bui et al., 2002] [Heinze, 2003] [Glover et al., 2004] [Schrempf and Hanebeck, 2005] due to the models' capability of the handling uncertainty present it human behaviour. Some of the most popular techniques which have been used in different recognition applications include Partially Observable Markov Decision Process (POMDP) [Hoey et al., 2007] [Taha et al., 2008] [Iba et al., 2003], Bayesian Networks (BN) [Jensen, 1996] [Song et al., 2010], Markov Networks [Castillo et al., 1997], Dynamic Bayesian Networks (DBN) and its variants [Tahboub, 2006] [Schrempf and Hanebeck, 2005] [Krauthausen and Hanebeck, 2009] and Hierarchical Hidden Markov Model (HHMM) [Liao et al., 2007] [Nguyen et al., 2005] [Zhu et al., 2008]. The selection of modelling technique depends on the intended outcome of the overall application. Figure 2.5 illustrates eight Markov process models, arranged in a cube the axes of which represents significant dimensions along which the models differ from each other [Mahadevan et al., 2004]. We will now discuss some of these probabilistic models which will be used in this thesis.

2.4.1 Bayesian Network

Bayesian Networks (BN) and their variants are probably the most widely employed belief network for modelling sequential data. BN constitute a more general and efficient way of expressing and computing with probability distributions. In a Bayesian network,

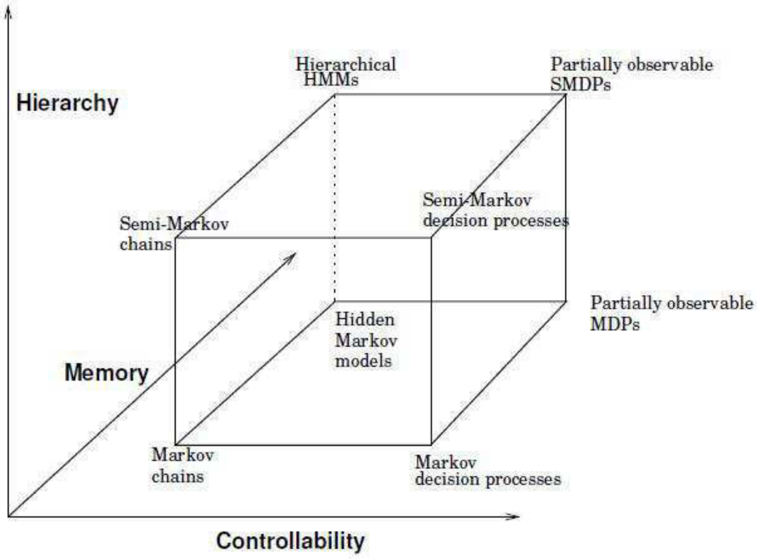


Figure 2.5: A spectrum of Markov process models along several dimensions: whether agents have a choice of action, whether states are observable or hidden and whether actions are unit-time (single step) or time-varying(multi-step) [Mahadevan et al., 2004]

it is possible to associate an arbitrary set of variables to the state of interest, and model the resulting joint probability distribution of the hidden state of interest. Using graphical model notation, BNs are represented by directed acyclic graphs (DAG) in which nodes represent variables and arcs/edges show the conditional dependencies among the variables. Each node on the net has an associated probability table, containing the conditional probabilities of the values that the node can take with respect to each of the possible combination of its parent nodes.

Assume that $x_1, ..., x_n$ are variables of a BN and $\pi_i (1 <= i <= n)$ is the set of parents of x_i. The joint probability distribution (JPD) of $x_1, ..., x_n$ can be factorised as:

$$Pr(x_1, ..., x_n) = \prod_{i=1}^{n} Pr(x_i|\pi_i) \tag{2.1}$$

2.4.2 Dynamic Bayesian Network

The Dynamic Bayesian Network (DBN) is a special Bayesian network, which adds a temporal dimension to single-slice BNs. The DBN consists of a sequence of time slices, and each time slice has a set of variables representing the state of the environment at a specific time. The temporal dimension makes the DBN capable of modelling sequential data. The evolution of the DBN model is represented by the links from the current time slice to the next time slice. The topology of a typical DBN is depicted in Figure 2.6.

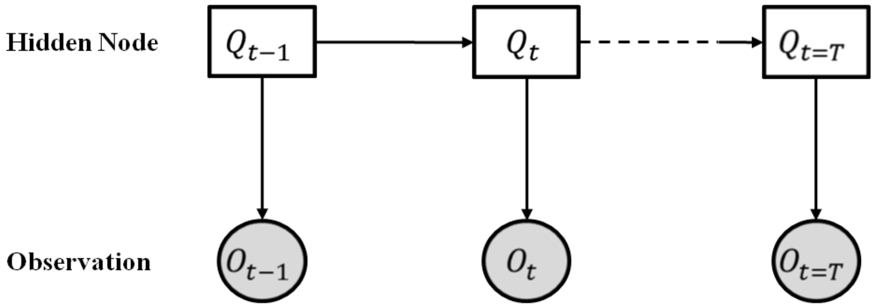


Figure 2.6: DBN model unrolled infinitely over an infinite time period

Representation

The structure of the DBN model comprises two types of nodes: Q_t and O_t. Edges between nodes represent their dependencies on each other. The hidden node Q_t corresponds to the state which is being inferred based on the observations perceived through node O_t. Observation nodes can be modelled as a mixture of Gaussian (μ, Σ) or discrete $P(O_t|Q_t)$ nodes based on the type of data. The topology of the connections between successive slices and between states and evidence variables in each time slice define the conditional dependencies between the variables. Furthermore, three clusters of information are needed to fully defined a DBN: the prior distributions $P(Q_t)$ when $t = 1$, the transition probability distribution between states $P(Q_t|Q_{t-1})$, and the conditional observation probabilities $P(O_t|Q_t)$. It should be noted that if Q_t represents a single scalar state then the DBN model becomes equivalent to a HMM model. The details of each of these probabilities are specified as below:

Prior Probability

The prior probabilities provides the initial probabilities or the most likely state of the hidden node. The initial probabilities are defined as:

$$P(Q_1) = \pi(j) \tag{2.2}$$

Transition Probability

The transition probability at time t, depends upon the previous state time step $t-1$ and defines transition from state i to j.

$$P(Q_t = j | Q_{t-1} = i) = A(i, j) \tag{2.3}$$

In Equation 2.2, π represents the initial probability whereas in 2.3, $A(i, j)$ represents the transition probability for state i to j.

Observation Probability

The observation model signifies the probability of seeing a specific observation conditioned on a discrete hidden state. The observation nodes can be modelled as both Gaussian and discrete. The CPDs for Gaussian and discrete nodes is given by:

$$\begin{gathered} P(O_t | Q_t = i) = N(\mu_i, \Sigma_i) \\ P(O_t | Q_t = i) = C(i) \end{gathered} \tag{2.4}$$

2.4.3 Hierarchical Hidden Markov Models

The HHMM framework used in our work is capable of structuring stochastic processes at multiple levels. The HHMM is an extension of the traditional HMM model, designed to model domains with a hierarchical structure including those with dependencies at

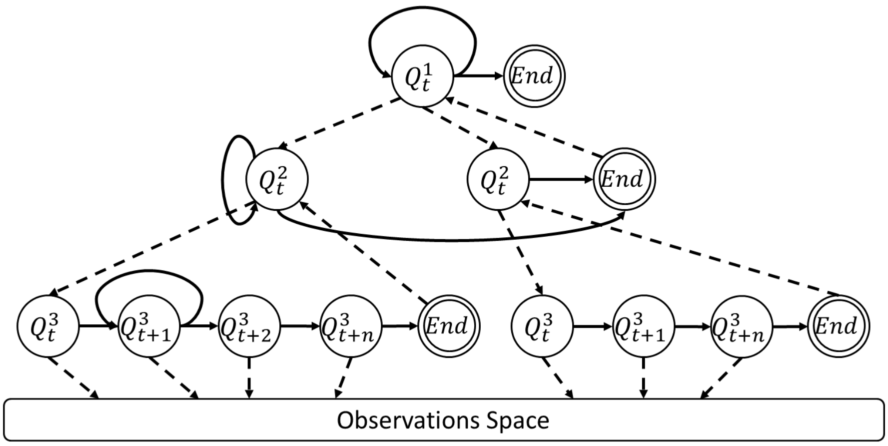


Figure 2.7: Example of a three level HHMM model where solid arcs represent horizontal transitions between states, and dotted arcs represent vertical transitions, i.e., connections between sub-HMMs. Double-ringed states represents end states (at least one per sub-HMM), where control flow is returned to the parent (calling) state. Each node at level 3 emits a single state based on the distribution over the observation space.

multiple length/time scales [Fine et al., 1998]. In a HHMM, the states of the stochastic automaton can emit single observations or strings of observations. Those that emit single observations are called "production states", and those that emit strings are termed "abstract states" [Murphy, 2002].

The example shown in Figure 2.7 provides an intuitive description of the process. The states at the highest level correspond to the abstract states, and are themselves governed by sub-HHMMs, entering into states Q^2. Since the states at level 2 being abstract, it enters its child HMM via its subsequence states Q^3. The horizontal transition in each child HMM (at level 3) emits a unique state with respect to the observations perceived by the model and is hence referred to as the "production state". Once the sub-HMM reaches the end state, the control is returned to the higher level, from wherever the sub-HMM sequence was called from. This is done recursively until the time when control is returned to the highest abstract state (level 1). The abstract state can transit to the next possible state only after all the sub-HMM at lower levels are terminated [Murphy, 2002].

The hierarchical nature of the HHMM model allows the decomposition of the problem at different levels of abstraction, thereby facilitating exploration (long term plan-

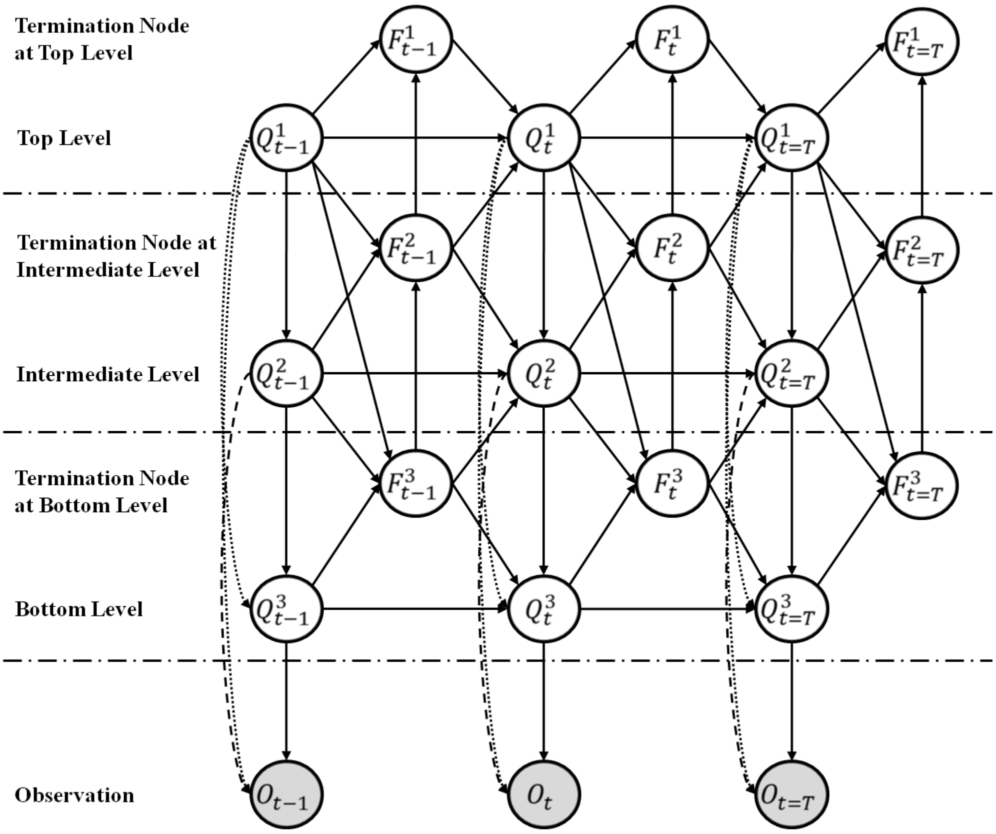


Figure 2.8: A 3-level HHMM model represented as Hierarchical-DBN Model. Q_t^d is the state at time t and level d; $F_t^d = 1$ if the sub-HMM at level d has finished (entered exit mode), otherwise $F_t^d = 0$. The nodes O_t are observed nodes which can be a mixture of both Gaussian and discrete nodes.

ning/activities) and exploitation (short term planning/action primitives) within the same framework. Within the paradigm of learning ADLs from APs, high-level activities (referred to as ADLs) call the more refined low-level activities (referred to as APs) according to some distribution. A low-level activity will in turn call another lower-level activity, and this process continues until the most primitive possible activity is performed. When the lower level activity terminates - in some state - the parent behaviour may also terminate as long as the current state is in the set of destination states of the parent node.

Representation

A HHMM framework can be represented as a Hierarchical Dynamic Bayesian Network (H-DBN) as shown in Figure 2.8. The structure consists of three types of nodes, Q_t^d, O_t, F_t^d where d is the depth of the hierarchy. The detail of each node is specified as follows:

- Q_t^d represents the state of the system at time t and level d. The state of the whole HHMM is encoded by the vector of $\vec{Q}_t = Q_t^1, Q_t^2, ..., Q_t^D$; intuitively, this encodes the contents of the stake, that specifies the complete "path" to take from the root(top most level state Q_t^1) to the leaf(bottom most level state Q_t^d) in the state transition diagram.

- The termination node F_t^d is a binary state indicator that is "on" if the sub-HMM at level d and time t has just finished the execution, otherwise it is "off". Note that if $F_t^d = 1$, then $F_t^{d^`} = 1$ for all $d^` > d$, which specifies which level we are currently on.

- As the true state of the user is hidden, observations node O_t are required which provides user/environment information. These nodes can be modelled as a mixture of Gaussian (μ, Σ) and/or as discrete $P(Q_t^d|O_t)$ node.

- The downward arc between the Q nodes represents the fact that a state "calls" a sub-state. The upward arcs between F nodes enforces the fact that higher level states in the HHMM model can only change states when the lower level ones are terminated (indicated by the he F node).

- Once the sub-HMM at lower level $d > 1$ terminates then the state at the top most level (Q_t^1) changes its state based on the initial probability at that level and observation probability.

Given the parameters (Q_t^d, O_t, F_t^d), the H-DBN defines the joint distribution over the set of variables that represents the evolution of the stochastic process over time. These distributions are in the form of prior distributions (initial probabilities), transition probabilities, termination probabilities and the observation probabilities. The prior distribution and the transition probabilities are defined at every level (d).

Prior Probability

The prior probability provides the initial probabilities of the most likely initial state. The initial probabilities at all levels are defined in Equation 2.5.

$$\begin{aligned} &P(Q_1^1 = j) = \pi^1(j) \\ &P(Q_1^d = j | Q_1^{1:d-1} = k) = \pi_k^d(j) \end{aligned} \tag{2.5}$$

where π^1 represent the initial probabilities for top level and π_k^d represents the same for $d = 2, ..., D$, given the state at higher levels is k.

Transition Probability

Each node in the H-DBN represents a conditional probability distribution (CPD) or table (CPT). Due to differences in the local topology and dependencies at different levels, we consider the bottom, middle/intermediate and top layers of the hierarchy.

The state at the top level at time t depends upon the previous state at the same level and the termination states at same level and intermediate level at time $t - 1$. Probabilities of the state Q_t^d at the highest level at time t is specified in Equation 2.6. The probabilities at the intermediate level d (where $d = 2 : D - 1$ & D is the maximum depth of the hierarchy) at time t depends on *(a)* the state at same level at time $t-1$, *(b)* the state of parent node $Q^{1:d-1}$ and *(c)* the termination node at the same level d and lower level $d - 1$. The termination node F^d specifies whether we should use transition probabilities or prior probabilities. The transition probabilities at the intermediate and at the bottom most ($d = D$) levels are defined in Equation 2.7 and Equation 2.8 respectively.

$$P(Q_t^1 = j | Q_{t-1}^1 = i, F_{t-1}^2 = b, F_{t-1}^1 = f) = \begin{cases} \delta(i,j) & \text{if } b = 0 \\ A^d(i,j) & \text{if } b = 1 \& f = 0 \\ \pi^d(j) & \text{if } b = 1 \& f = 1 \end{cases} \tag{2.6}$$

$\delta(i,j)$ is the Kronecker delta function which specifies the self transition in the same state given the termination state at the lower level is "off". A^d is the transition probability from state i to j at the top level given the termination state at the lower level is "on" and at the same level is "off". Similarly, π^d is the initial distribution at the top level given the termination state at both the lower and same level is "on".

$$P(Q_t^d = j | Q_{t-1}^d = i, F_{t-1}^{d+1} = b, F_{t-1}^d = f, Q_t^{1:d-1} = k) = \begin{cases} \delta(i,j) & \text{if } b = 0 \\ A_k^d(i,j) & \text{if } b = 1 \& f = 0 \\ \pi_k^d(j) & \text{if } b = 1 \& f = 1 \end{cases} \tag{2.7}$$

$\delta(i,j)$ is the transition probability from state i to j at level d when the termination state at the lower level is "off". A_k^d is the transition probability from state i to j at level d given the termination state at the lower level $(d-1)$ is "on" and at the same level is "off" and the parent nodes are in state k. Similarly, π_k^d is the initial distribution at level d given the termination state at both the lower $(d-1)$ and same level is "on" and the higher level state is k.

$$P(Q_t^D = j | Q_{t-1}^D = i, F_{t-1}^D = f, Q_t^{1:D-1} = k) = \begin{cases} A_k^D(i,j) & \text{if } f = 0 \\ \pi_k^D(j) & \text{if } f = 1 \end{cases} \tag{2.8}$$

A_k^D is the transition probability from state i to j at level D (the bottom most level) given the termination state at the same level is "off" and the parent node is in state k, π_k^D in Equation 2.8 is the initial distribution at level D given that the parent node is in state k. In Equation 2.6, 2.7 and 2.8, we assume that $i, j \neq \textit{end state}$.

Termination Model

The termination probabilities at top, intermediate and bottom levels will be different due to the topological differences in the structure of the H-DBN model. The termination probabilities at these three levels are given by Equation 2.9, 2.10 and 2.11 respectively.

$$P(F_t^1 = 1 | Q_t^d = i, Q_t^{1:d-1} = k, F_t^2 = b) = \begin{cases} 0 & \text{if } b = 0 \\ A^1(i, end) & \text{if } b = 1 \end{cases} \tag{2.9}$$

$$P(F_t^d = 1 | Q_t^d = i, Q_t^{1:d-1} = k, F_t^{d+1} = b) = \begin{cases} 0 & \text{if } b = 0 \\ A_k^d(i, end) & \text{if } b = 1 \end{cases} \tag{2.10}$$

$$P(F_t^D = 1 | Q_t^{1:D-1} = k, Q_t^D = i) = A_k^D(i, end) \tag{2.11}$$

A^1, A_k^d and A_k^d are the termination states at top, intermediate and bottom level.

Observation Probability

The observation probability for a H-DBN model is similar to that defined in Equation 2.4, which is conditioned on seeing a specific observation on a discrete hidden state.

Advantages of H-DBN Model over DBN Model

There are several advantages of the H-DBN model over a flat DBN model. Due to its hierarchical nature, the H-DBN model makes multi scale interpretation of data possible which is otherwise not possible with a flat DBN model. The hierarchical nature also provides the advantage of modularity whereby the same sub-HMM model of the H-DBN model can be reused for different state space. The H-DBN also exploits any constraints that may be present in the parameters which cannot be exploited efficiently by the DBN model due to its flatness.

2.4.4 Learning

Various learning techniques both supervised and unsupervised can be used for learning the DBN and H-DBN model. Expectation Maximisation (EM) [Blimes, 1998] and its variants is one of the most popular statistical techniques used for unsupervised learning. In this work we used EM for learning the model and a maximum likelihood estimator for inference. The EM algorithm optimises the model parameters to better fit the training data. It performs this by iterating between an Expectation step (E-step) and Maximization step (M-step). In each E-Step it estimates the expectations (distributions) over the hidden variables using the observations along with the conditional probability density (CPD) of the model. Then in the M-step the model parameters (i.e. the CPD's) are updated using the expectations of the hidden variables obtained in the E-step. Each iteration would continue to improve the estimates of the hidden variables and will eventually converge to a local optimum. The EM algorithm for learning the parameters in the DBN and HDBN models is given in Algorithm 1. It should be noted that the model parameters (initial π, transition $A(i, j)$ and observation $B(i, k)$ probabilities) are manually initialised based on common laws of operation and acceptable user behaviour which then are further optimised by the EM algorithm.

2.4.5 Inference

The aim of the inference algorithm is to calculate the marginals $P(Q_t = i|O_{1:\tau})$ at the next time slice t given the sequence of observations and maximise them. If $\tau = t$, i.e. the observation sequence is available up to the current time, then the inference is called filtering, also known as on-line inference. On the other hand, when $\tau = T$ then the inference of state Q at time t is done using the entire sequence of observation, more often referred to as smoothing or off-line inference.

For inference of ADLs and APs we utilised Viterbi algorithm (max-product algorithm) for off-line inference, which maximises the sequence of states Q from $t = (0 : T)$ given the entire observation sequence $O_{t=0:T}$. For on-line inference we utilised the forward algorithm which performs inference of state Q at time t given the observation sequence

Algorithm 1 :The Expectation Maximisation (EM) algorithm to learn the parameters in DBN and HDBN Models [Murphy, 2012]

Input:

Observation sequences $O_1, O_2, O_3, ..., O_T$

Output:

Model Parameter $\theta = (\pi, A(i,j), B(i,k))$

Initial State Probability π

State Transition Probability $A(i,j)$

Observation Probability $B(i,k)$

Begin

Initialise $\pi^{(t=1)}, A(i,j)^{(t=1)}, B(i,k)^{(t=1)}$

for $x = 1 \rightarrow N$ **do** ▷ /* N is number of iteration for convergence */

E-Step: compute expected value using old (θ^{old}) model parameter

From $\pi^{(x)}, A(i,j)^{(x)}, B(i,k)^{(x)}$ and $O_1, O_2, O_3, ..., O_T$ compute:

Auxiliary function:

$$F^n(\theta, \theta^{old}) = \sum_{i=1}^{N} E[N_i^1] \log \pi_i + \sum_{i=1}^{N}\sum_{j=1}^{N} E[N_{ij}] \log A_{ij} + \sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{k=1}^{K} P(q_t = k | O_i \theta^{old}) \log p(O_t = i | \phi)$$

where:

Expected count of initial distribution $\pi^{(x)}$ is given by

$$E[N_i^1] = \sum_{i=1}^{N} p(q_1 = i | O_i, \theta^{old})$$

Expected count of transition distribution $A(i,j)^{(x)}$ is given by

$$E[N_{ij}] = \sum_{i=1}^{N}\sum_{t=2}^{T_i} P(q_t = j, q_{t-1} = i | O_i, \theta^{old})$$

Expected count of observation distribution $B(i,k)^{(x)}$ given by

$$E[N_j] = \sum_{i=1}^{N}\sum_{t=1}^{T_i} P(q_t = j | O_i = k, \theta^{old})$$

M-Step: Normalise expected counts of model parameter $\pi^{(x)}, A(i,j)^{(x)}, B(i,k)^{(x)}$

$$\pi^{(x)} = normalise \sum_{O_1}^{O_T} \pi^{(x)}$$

$$A(i,j)^{(x)} = normalise \sum_{O_1}^{O_T} A(i,j)^{(x)}$$

$$B(i,k)^{(x)} = normalise \sum_{O_1}^{O_T} B(i,k)^{(x)}$$

end for

return Model Parameter $\theta = (\pi, A(i,j), B(i,k))$

End Begin

till time t. The algorithms for off-line inference and on-line inference are given in Algorithm 2 and Algorithm 3 respectively.

Algorithm 2 :Viterbi Algorithm to estimate the hidden state [Murphy, 2012]

Input:

Observation sequences $O_1, O_2, O_3, ..., O_T$

Initial State Probability $P(Q_1 = i)$

State Transition Probability $P(Q_t = j|Q_{t-1} = i)$

Observation Probability $P(O_t = o|Q_t = j)$

Output:

Estimate hidden state $Q_1, Q_2, Q_3, ..., Q_T$

Begin ▷ /* recursively estimate Q_t till $t = 1$ */

$$Q_{1:T} = \underset{Q_{1:T}}{\operatorname{argmax}} P(Q_{1:T}|O_{1:T})$$

$$= \underset{Q_{1:T}}{\operatorname{argmax}} P(Q_{1:T}, O_{1:T})$$

$$= \underset{Q_{1:T}}{\operatorname{argmax}} \Big(P(Q_1)P(O_1|Q_1)\prod_{t=2}^{T} P(Q_t|Q_{t-1})P(O_t|Q_t)\Big)$$

$$= \underset{Q_{t=T}}{\operatorname{argmax}} \Big(P(O_{t=T}|Q_{t=T})\underset{Q_{1:T-1}}{\operatorname{argmax}}(P(Q_1)P(O_1|Q_1)\prod_{t=2}^{T-1} P(Q_t|Q_{t-1})P(O_t|Q_t)* P(Q_T|Q_{T-1}))\Big)$$

$$= \underset{Q_{t=T}}{\operatorname{argmax}} \Big(P(O_{t=T}|Q_{t=T})\underset{Q_{T-1}}{\operatorname{argmax}}(P(O_{t=T-1}|Q_{t=T-1})P(Q_T|Q_{T-1})... \underset{Q_{t=1}}{\operatorname{argmax}} P(Q_1)P(O_1|Q_1)P(Q_2|Q_1))\Big)$$

End Begin

2.5 Summary

In the first part of this chapter we reviewed the importance of ADLs and the challenges associated with modelling these activities. We described our approach of how to develop a dictionary of action primitives which best describe the decomposition of complex ADLs into atomic actions. The proposed architecture for representing ADLs using APs also reduces the entire search space to a cluster of APs which can be combined in different sequence to construct any complex ADL.

In the second part of this chapter, we detailed various Bayesian networks based prob-

Algorithm 3 :On-line inference performed using only the forward part of the forward-backward Algorithm to estimate the hidden state [Murphy, 2012]

Input:

Observation sequences $O_1, O_2, O_3, ..., O_t$

Initial State Probability $P(Q_1)$

State Transition Probability $P(Q_t = j|Q_{t-1} = i)$

Observation Probability $P(O_t = o|Q_t = j)$

Output:

Estimate hidden state $Q_1, Q_2, Q_3, ..., Q_t$

Begin ▷ /* O_t is the observation sequence till current time t*/

$Q_t = P(Q_t|O_{1:t})$

$= P(O_t|Q_t, O_{1:t-1})P(Q_t|O_{1:t})$

$= P(O_t|Q_t)(\sum_{Q_{t-1}} P(Q_t|Q_{t-1})P(Q_{t-1}|O_{1:t-1}))$

End Begin ▷ /* recursively estimate Q till $t = 1 * /$

abilistic models which are widely used in human activity recognition. Both DBN & HHMM based probabilistic frameworks proved to be an ideal tool to model activities which have previous time dependencies. With its ability of modelling activities at multiple levels of the hierarchy, the HHMM framework provides the necessary tools to model high level user activities from low level sensor measurement. In the next chapter, we address the problem of modelling ADLs which are performed by a user of a mobility aid device. The pool of activities consists of navigational (*visiting locations of interest*) and support(*stand-up or sit-down*). We apply our approach of constructing a dictionary of meaningful APs which are combined in different sequence to model ADLs.

Chapter 3

Modelling Activities of Daily Living using Topological Maps

3.1 Introduction

In the previous chapter we introduced an efficient strategy to model Activities of Daily Living (ADLs) by decomposing them into a sequence of action primitives (APs). In this chapter we focus on developing a framework for modelling everyday activities performed using the support of a power walker as a mobility device. The pool of activities performed using this device can be navigational (visiting locations of interest) and/or support (using the support of the walker to stand up or sit down) activities. Probabilistic models as described in Chapter 2 are used to capture the complexity and ambiguity present in human behaviour while he/she performs different everyday activities.

A variety of ADLs are performed by a user, seeking support of a walking mobility device. These activities are a combination of support/static and navigational activities which is a natural way in which, user performs most of the ADLs. For instance, for a user to navigate to the bathroom, he/she has to perform a sequence of activities such as standing up using the support of the walker *(support activity)*, followed by walking towards the bathroom *(navigational activity)*. ADLs performed by walker users are a sequential combination of support and navigational activities. The sequential dependencies between support and navigational activities provides important information related to patterns in which these ADLs are performed. The sequential relationship between different activities cannot be exploited if ADLs are modelled in an isolated manner. The coupling of both these activities is tackled using a state of the art probabilistic method. This technique is based on a hierarchical probabilistic model where both type of activities are inferred using a single unified model that embeds both user's navigational desire to navigate to a particular location of interest and support activities such as standing up or sitting down using the support of the walker. For completeness we also compared the inference accuracy of the hierarchical approach with a more conventional probabilistic models such as a Dynamic Bayesian Network (DBN) and a combination of a hierarchical probabilistic model and a discriminative classifier.

3.2 Related Work

Recognising different ADLs carried out by a mobility aid device user is not an uncommon research area. Based on the targeted problem, several methodologies have been applied by different research group. However, due to the demanding nature of the solution, almost all such methodologies have shortcomings in one way or other. The following section describes the attempts that have been made in the area of recognising ADLs in general and ADLs for mobility device users specifically.

Conventional walking aid devices or power wheelchairs have been deployed successfully to support the performance of a large array of ADLs. Researchers have worked on developing intelligent mobility aid devices designed to give a sense of safety and security to the user during the course of performing various ADLs. Intelligent/Smart walkers have recently begun to emerge as an alternative assistive device. The very first walker based intelligent mobility device was developed by Lacy [Gerard and Kenneth, 1998]. Under this project they developed a personal adaptive mobility aid(PAM-AID), which physically supports a walking user and provides obstacle avoidance to ensure safer travel. In 2003, Morris and colleagues [Morris et al., 2003], developed a smart walker platform which not only provided basic assistance in terms of safety while navigating in a cluttered environment but also provided assistance with navigation and global orientation. Both of these intelligent walking frames developed by Lacy and Morris provided the user with safe navigational support. Wasson *et. al.* [Wasson et al., 2004] developed a COOL-Aide smart walker to operate in a more tightly coupled, shared control loop with its human user. Instead of active guidance, the COOL-Aide provides a passive shared control system that delivers active steering assistance only as needed, and no propulsion assistance. The main purpose of this system is to derive the navigational intention of the user based on measuring forces and moments applied to the walker's handles and to provide steering assistance as and when required. Omar *et. al.* [Omar et al., 2010] proposed an activity recognition technique based on Hidden Markov Model (HMM) and Conditional Random Fields (CRFs) for an instrumented passive rollator walker. The model recognized a number of user states: not touching the walker, stop/standing, walking forward, turn left, turn right, walking backwards and transfers (sit to stand/stand to sit). A Hierarchical Semi-Markov Model (HSMM) was proposed by Glover *et. al.* [Glover et al., 2004] to learn the user's walking activities. The walking activities inferred by the HSMM framework were defined based on the topological region visited by the user in a given environment. The HSMM framework operated at three different levels: at the lowest level the metric motion is described by metric coordinates, at the mid-level the framework uses topological regions as its element and at the highest level the person's walking activities are divided into logically broader walking activities.

Despite the impressive research outcome accomplished in the reviewed literature for activity recognition of an assistive device user, the research did not address the problem of predicting comprehensive ADLs which consists of both support and navigational ADLs. This becomes an important criterion, as such a system would be capable of predicting and assisting in the overall ADLs performed by the device user.

In this chapter we propose a system that adds the ability to integrate both support activities and navigational activities within a single probabilistic model. This is achieved by decomposing ADLs into smaller sub-activities called APs. For a mobility device user most of the ADLs consists of visiting different location of interest which are complemented with some support activities. Navigational ADLs are further decomposed into intermediate junction points. Junction points are intermediate locations in a given environment from where two or more possible user paths exists. These intermediate junction points are mapped to APs which are combined in different sequence to represent the overall navigational activity a user is trying to perform. We deploy a topological representation of map (as shown in Figure 3.1) to encode the environment, as they offer a compact structure that better matches the human's natural description of a path (e.g."take a left at the second junction" instead of "go straight for 100 meters, then turn left") [Rawlinson and Jarvis, 2008]. Different techniques currently exists on how to construct topological maps, such as imprecise human drawings [Setalaphruk et al., 2003] and generalised Voronai diagrams [Liao et al., 2003] [Aurenhammer, 1991].

The inference of ADLs, relies on both human behaviour of the user as perceived by the low level sensors fitted on the assistive device and the location of the user in a given environment. This differs from the work presented by Wasson *et. al.* [Wasson et al., 2004], Glover *et. al.* [Glover et al., 2004] and Omar *et. al.* [Omar et al., 2010], where the system provides either navigational assistance to reach the respective goal location or assistance in performing static activities such as standing up using the support of the walker, but not both. The intelligent framework proposed in this work applies the strategy of decomposing ADLs at multiple levels. The user is always in control yet the system is designed to actively yet unobtrusively assist the users as they go about their normal daily activities without the need for any explicit actions such as pushing buttons/panels or voice commands. This is an important motivation for this work, since it is assumed that the intended user population may not always be cognitively or physically able of pro-actively providing such unequivocal indications of their needs. Further the inference of high level ADLs, allows the system to provide assistive support to the user population suffering from different cognitive disability such as Alzheimer, dementia etc. where users tend to have memory lapses and forget the ADLs they intend to perform. In such scenarios inferring high level ADLs becomes intuitively important as the system would be capable of providing support and assist the user in achieving the task that they were intending to perform.

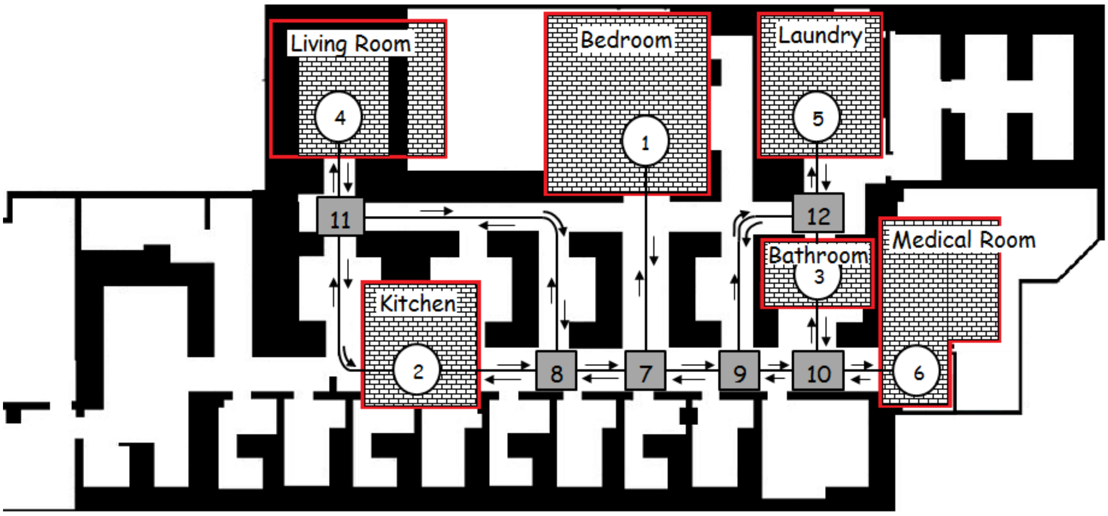


Figure 3.1: A topological map structure of an office area representing a typical home setting is shown on top of the metric representations. Circle nodes in this structure represents a possible location of interest, rectangle nodes represents junction points in the topological map and line segments represent a viable physical path between two nodes.

3.3 Capturing Navigational and Support Activities

Capturing navigational activity is the process of acquiring knowledge about human activities by monitoring how people go about achieving their routine ADLs. Typically, people requiring mobility assistance in a constrained environment such as a home or a elderly home care, have a known set of target locations that they visit to perform their daily tasks, such as *bathroom, bedroom, living room, medical facility, etc.* Support activities are another set of activities which are often performed along side the navigational activities. In terms of using walker as a support device, activities such as *standing up from sitting* or *navigating locally in a room* are termed as static/non-navigational activities. Capturing human behaviour while performing different ADLs provides important parameter information as it enables the intelligent system to learn these complex behaviour and to predict how, when and where a user is in need of assistance.

For capturing user behaviour while they perform navigational activities, the spatial information of the environment is represented using a structural topological map. A topological graph representation of the environment consists of vertices which represents locations that a user would visit, junctions points which represents one or more possible paths from a given location and edges representing a viable path connecting two locations through various junction points. For cluttered indoor environments, the map topology can be represented by a graphical tree of nodes and connections (segments), where a set of nodes represents a location of interest or junction points in the map and the connection represents a physical path that connects two locations. The topological map used in this work is shown in Figure 3.1. The map is the structure of an office area representing a typical home setting. The topological map was generated before hand by dividing the map using 12 junction points and segments interconnecting these junction points. As shown in Figure 3.1, the junction points were either the end destinations or were designated to a point from where there were more than two possible paths to the next junction.

3.3.1 Locations of Interest

Locations of interest can be thought of as spatial locations in the environment where a human spends the majority of his/her time and these locations vary depending on the user and their environment. Identifying locations of interest is not an easy task and there is no general method to achieve this, as it is a user and environment dependent problem. This information can be manually specified or extracted by monitoring user patterns which is usually done with the help of an occupational therapist [Practice, 2008]. In general, remaining at a certain location for more than a predefined time can be used to identify locations of interest. Knowing a set of possible destinations is necessary for predictive robot navigation where the exact destination is not known but predicted based on a sequence of observations from a human user and/or from the environment. For the current project, we defined 6 locations of interest (Figure 3.1) which a typical walker user will visit to perform various ADLs.

3.4 Dictionary of Action Primitives

The complex activities performed by a walker user are normally a combination of both navigational and support activities. The navigational activities can further be categorised into goal oriented navigation and manoeuvring locally in a room. The dictionary of Action Primitives (APs) used in this work consists of different segments obtained through the topological representation generated by connecting the vertices. For example the navigational route of visiting the kitchen from the bedroom would consist of going through segments '*1 to 7, 7 to 8 & 8 to 2*'. Different segments of an entire navigational activity, which acts as APs, are defined such that it can be directly mapped to the control system of the robot. This enables the robot to assist the user to achieve their intended goal which involves manoeuvring through different segments.

On the other hand support activities such as *standup*, *sitdown*, *recalling the walker* etc., cannot be decomposed any further due to two reasons. Firstly, due to the limitation of the low level sensor, the number of key user poses cannot be efficiently perceived by the sensors and secondly, the pool of APs were generated such that they can directly be mapped to the control of the walker so as to provide necessary assistance, hence modelling key user poses did not add any valuable information for providing the required assistance.

3.5 Inferring ADLs using Probabilistic Models

Stochastic or probabilistic models, also known as belief networks, encompass a wide range of generative algorithms particularly suited to applications where variables evolve over time and there is incomplete knowledge or randomness in the processes involved. The model consists of a sequence of time-slices where each time slice is described by a set of random variables which represent the state of the user at the current time [Tahboub, 2006], and its dependencies on previous state(s). In general, the key advantage of using stochastic models as opposed to, for instance, deterministic models described by rigid mathematical laws, is their ability to capture complex dynamics that are not completely observable or unambiguous. The added advantage of being able to represent these relationships using graphical models [Jensen, 1996] has also made them a popular tool, particularly with the Artificial Intelligent (AI) community.

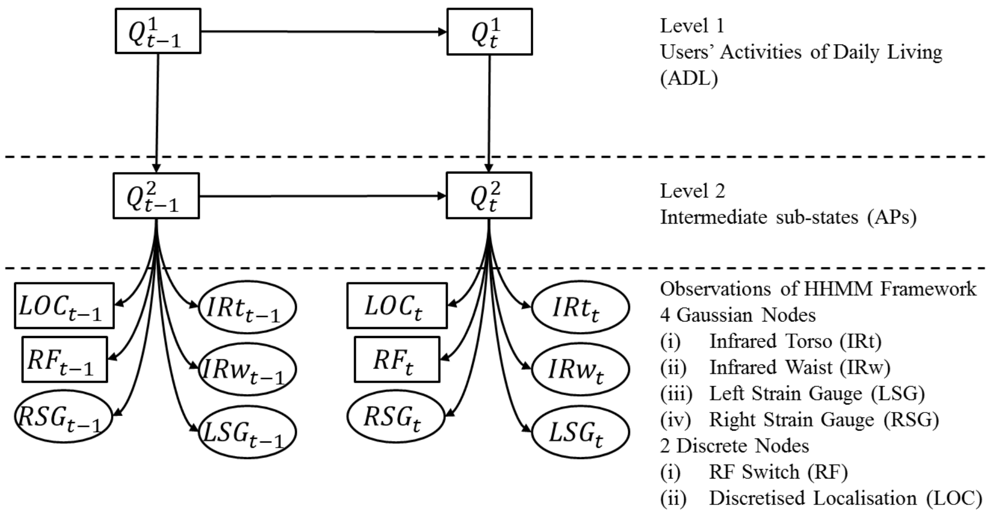


Figure 3.2: Layered DBN model used to infer ADLs performed by walker user

3.5.1 Layered Dynamic Bayesian Network (L-DBN) to model ADLs

As specified in Chapter 2, DBN is a variant of Bayesian Network (BN) which adds a temporal dimension to single-slice BN. It consists of a sequence of time slices, with each time slices having a set of variables representing the state of the environment at a specific time. The L-DBN model used to infer ADLs is shown in Figure 3.2. The model consists of two layers, where the ADLs and the corresponding APs are modelled at different levels. The hidden nodes of the model are denoted by Q_t which corresponds to the state of the user whereas the observation nodes represent the sensing of the physical variables and the environment. The edges between the nodes represents the conditional dependencies between user states and observations as perceived by the sensors. Apart from a graphical model the DBN model needs to be provided with an initial belief in the form of prior probabilities $(P(Q^1_{t=1}), P(Q^2_{t=1}))$, transition probabilities $(P(Q^1_t|Q^1_{(t-1)}), P(Q^2_t|Q^2_{t-1}, Q^1_t))$ and observation probabilities $(P(O_t|Q^2_t))$. These probabilities are further optimised from the data during the learning phase.

3.5.2 Hierarchical Hidden Markov Model (HHMM) to model ADLs

Hierarchical Hidden Markov Model (HHMM) are used to structure multi-level stochastic processes. In this application, ADLs are hierarchically split-up at different levels of

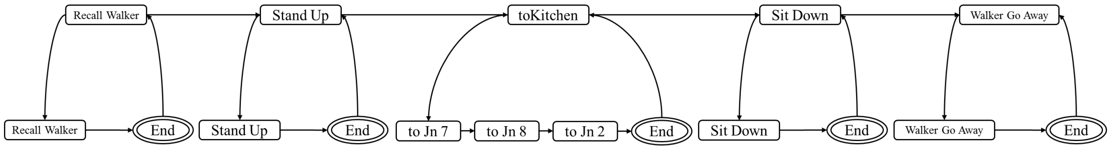


Figure 3.3: Time series of activities performed in sequence and their further decomposition into sequence of APs

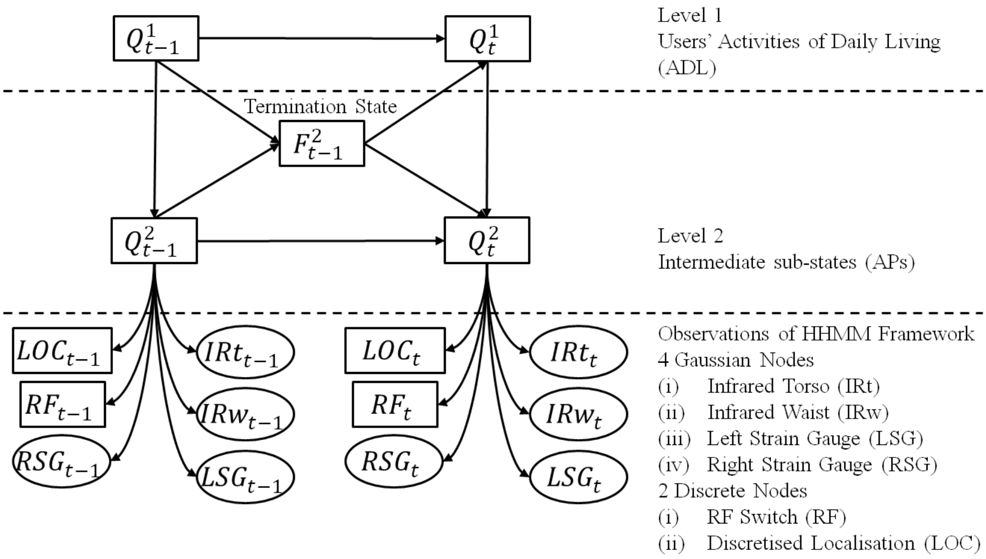


Figure 3.4: Hierarchical-DBN representation of a 2-level HHMM Framework used to infer ADLs. The horizontal dashed lines indicate levels of hierarchy

abstraction. User ADLs are represented at the top level while the intermediate level (Level 2 in Figure 3.4) represents the sequence of APs into which the ADLs can be decomposed. Walker assisted ADLs can be divided into support activities which are static in nature and are generally associated with being performed at a single location (*e.g. stand up, sit down, recall mobility device*) and secondly, navigational ADLs which involve visiting location of interest (*e.g. going to kitchen*). As stated in Section 3.4, support activities cannot be decomposed further and hence such ADLs are also regarded as independent APs in themselves. On the other hand, navigational activities can be further decomposed into more basic navigational cues, for instance, walking toward a junction shown in the topological map in Figure 3.1.

A typical scenario could be illustrated by considering for instance the activities involved in going to the kitchen from the bedroom (shown in light blue in Figure 3.6). The sequence of activities can be decomposed as "*Recall Walker* ⇒ *Stand Up* ⇒ *Navigate in Room* (until out onto the corridor) ⇒ *To Kitchen* [go to Junction 7 ⇒ go to Junction 8 (until entering kitchen)] ⇒ *Navigate in Room* ⇒ *Sit Down* ⇒ *Walker Go Away*" (The decomposition of the activity into navigational and support ADLs and their respective transitions from one ADL to another is shown in Figure 3.3)

The HHMM framework used to model ADLs can be represented as a Hierarchical Dynamic Bayesian Network (H-DBN) as shown in Figure 3.4. The framework consists of a 2-level hierarchy (d = 2) and consists of three types of nodes, Q_t^d, O_t and F_t^d where $d = 2$ is the depth of the hierarchical structure. The intermediate level states are the APs which are inferred from the physical sensor whereas the ADLs are inferred from the state sequence of APs. The node Q_t^1 corresponds to ADLs that the user is performing whereas the Q_t^2 corresponds to the respective APs. It should be noted that the decision to use a 2-level hierarchical structure was based on the fact that the pool of APs generated at the 2^{nd} level were simple enough to be inferred directly from the sensor information. The observation nodes are a combination of both physical sensor installed on the walker platform and localization information. The details of observation each observation sensors are as follows:

- 4 Gaussian nodes are the readings from the physical sensors installed on the walker (IRt, IRw, LSG, RSG).
- 2 discrete nodes, RF and Localization. An RF switch is used by the user to direct the walker to go away and recall it as needed. Location is derived from a localiser in a topological manner: hence, instead of using the metric (x, y) information provided by the localiser, 19 discretised locations are supplied as an observation to the HHMM network. The 19 discretised locations were generated beforehand using grid tiles of different sizes. The size of the grid was decided in such a way that the number of representative locations used in the experiments were reduced for computational purposes, yet there was no loss of topographical information for the navigational ADLs. Further each of these 19 discretised location corresponded to the location of the segment joining the respective junction points shown in Figure 3.1. Our office environment (shown in Figure 3.1 was considered as a

representation of a typical home environment, and the geometrical space was divided into the relevant points of interest a user would normally visit during the day.

Given the parameters (Q_t^d, O_t, F_t^d), the H-DBN defines the joint distribution over the set of variables that represents the evolution of the stochastic process over time. These distributions are in the form of prior probabilities (initial probabilities of the state variables at each level), transition probabilities, termination probabilities and observation probabilities. The prior probabilities at both levels are defined by Equation 3.1, while the transition probabilities at Level 1 and Level 2 are given by Equation 3.2 and 3.3 respectively. The termination probability is given by Equation 3.4. The observation nodes are modelled as both Gaussian and discrete. The CPDs for Gaussian and discrete nodes is given by Equation 3.5. Within the paradigm of modelling walker assisted ADLs, the initial location from where the user starts his/her ADL is incorporated by the initial probabilities of the H-DBN and L-DBN model, whereas the end location are modelled by the termination probabilities for the HHMM model. The sequence of APs will vary based on the start and end location, hence selection of the appropriate AP and

Figure 3.5: Power Walker used as a Mobility Device

the transition between APs is modelled via the transition probabilities between APs at level 2 of the H-DBN and the L-DBN model.

$$\begin{aligned} P(Q_1^1) &= \pi^1(j) \\ P(Q_1^2) &= \pi_k^2(j) \end{aligned} \tag{3.1}$$

$$P(Q_t^1 = j | Q_{t-1}^1 = i, F_{t-1}^2 = f) = \begin{cases} A^1(i,j) & \text{if } F_{t-1}^2 = 0 \\ \pi^1(j) & \text{if } F_{t-1}^2 = 1 \end{cases} \tag{3.2}$$

$$P(Q_t^2 = j | Q_{t-1}^2 = i, F_{t-1}^2 = f, Q_t^1 = k) = \begin{cases} A_k^2(i,j) & \text{if } F_{t-1}^2 = 0 \\ \pi_k^2(j) & \text{if } F_{t-1}^2 = 1 \end{cases} \tag{3.3}$$

$$P(F_t^2 = 1 | Q_t^2 = k, Q_t^1 = i) = A_k^2(i, end) \tag{3.4}$$

In Equation 3.1, 3.3, 3.2 and 3.4, A^1 and π^1 represent the transition and initial probabilities respectively at level 1 where as A_k^2 and π_k^2 represents the same at Level 2 given the state at Level 1 is k.

$$\begin{aligned} P(O_t | Q_t^1 = i) &= N(\mu_i, \Sigma_i) \\ P(O_t | Q_t^1 = i) &= C(i) \end{aligned} \tag{3.5}$$

3.6 Power Walker for Data Collection

Data were collected using a power walker (Figure 3.5) which was used as an aid device to perform various ADLs. The low level hardware sensor consists of two proximity sensors, four strain gauges (2 on each handle bar), a RF switch and a Hokuyo URG-04LX laser range finder. The frame was also instrumented with two gear-head motors and incremental optical encoders installed at the rear wheels. Further hardware and software details of the platform can be found in Appendix A and C respectively.

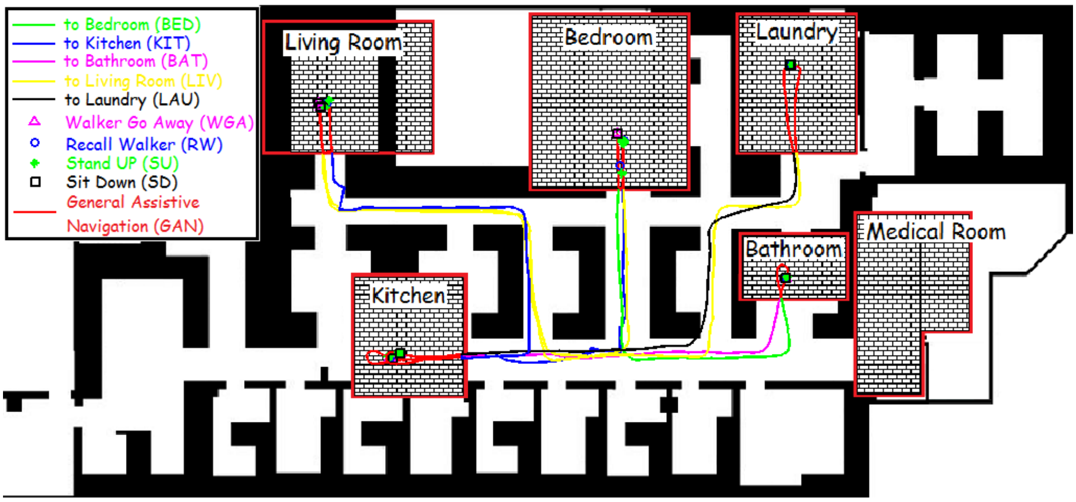


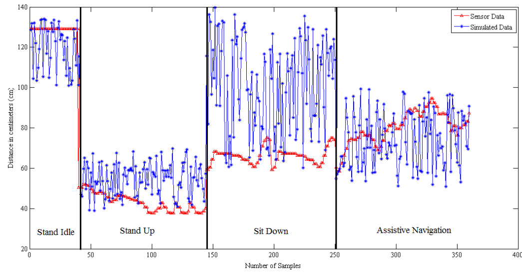
Figure 3.6: 2-D Bird's eye view of environment divided into typical locations of interest in an home, superimposed with different activity routines performed by one of the walker user

3.7 Results

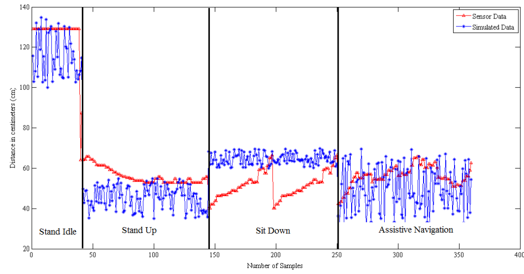
To evaluate our proposed methodology of modelling ADLs by exploiting the ADLs-APs relationship and the hierarchical characteristics of the HHMM model we tested the model with both synthetic data and real-time data collected using the power walker described in Appendix A. The HHMM framework was tested off-line using both datasets. We used the BNT toolbox [Murphy, 2002] (a popular and versatile toolbox for modelling different graphical models) to learn and infer ADLs using both layered DBN (L-DBN) and the HHMM frameworks. Unsupervised learning in the form of Expectation Maximization was used to learn ADLs, and the Maximum Likelihood Estimator was utilised for inference.

3.7.1 Evaluation using Synthetic Data

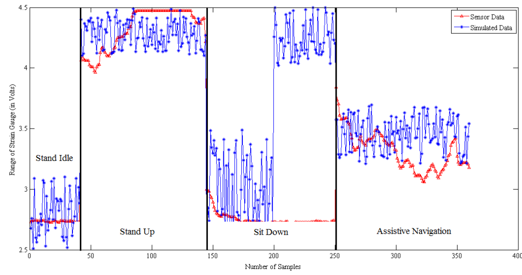
Large training and testing data sets were required for extensive evaluation and meaningful analysis of the proposed strategy. Hence, synthetic data for each sensor fitted on the walker was generated which involved visiting all the location of interest from every other location (i.e. bedroom visited from every other location of interest in the environment as shown in Figure 3.1). The sensor ranges and their expected behaviours



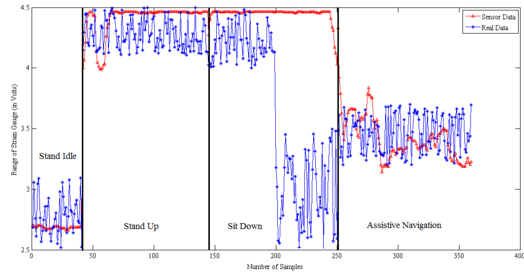
(a) Infra Red Torso (IRt) sensor



(b) Infra Red Waist (IRw) sensor



(c) Left Strain Gauge (LSG) sensor



(d) Right Strain Gauge (RSG) sensor

Figure 3.7: The data plot shows a comparison between real sensor data and synthetic data generated while the user performs four basic primitive actions which are combined with other sensors to model the overall ADLs.

were studied beforehand by performing smaller scale experiments on the walker platform with different able users to perceive their actions/behaviours. The synthetic data generated for all the sensors were analogous to that produced by real sensors. The two infra-red sensors (IRt, IRw) and the strain gauges (LSG, RSG) produced continuous data whereas the RF switch being an on/off switch produced discrete data. Further the localisation data produced by the Monte Carlo localiser particle filter using the laser and wheel encoder readings were continuous in nature and hence similar data was generated $(x, y, orientation)$ synthetically. For example, the infra-red sensor and the strain gauges had specific behaviour as shown in Figure 3.7 for different ADLs and hence the synthetic data generated for ADLs had similar characteristics.

Further the synthetic data generated for different ADLs and their corresponding APs had to be continuous over a time period so as to capture the entire user behaviour similar to the data collected from the real sensors while the user performs different ADLs. To illustrate, the sequence of activities that a user would perform to go to the living room from the bedroom, would be recalling the walker (**RW**), followed by standing up (**SU**)

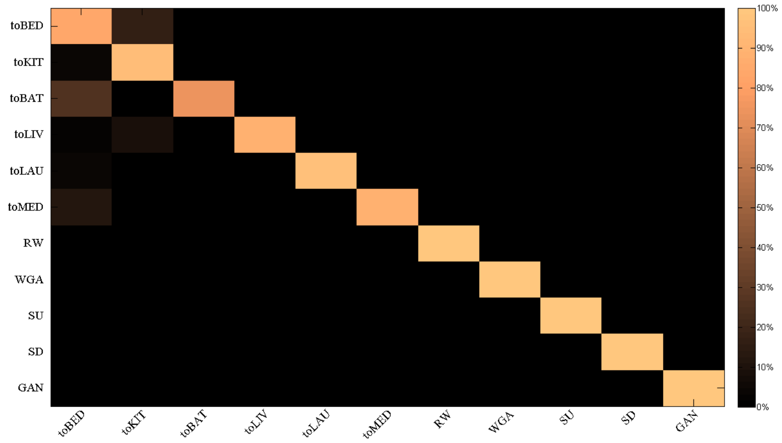


Figure 3.8: Confusion matrix for ADLs inferred by HHMM Model at level 1 using synthetic data

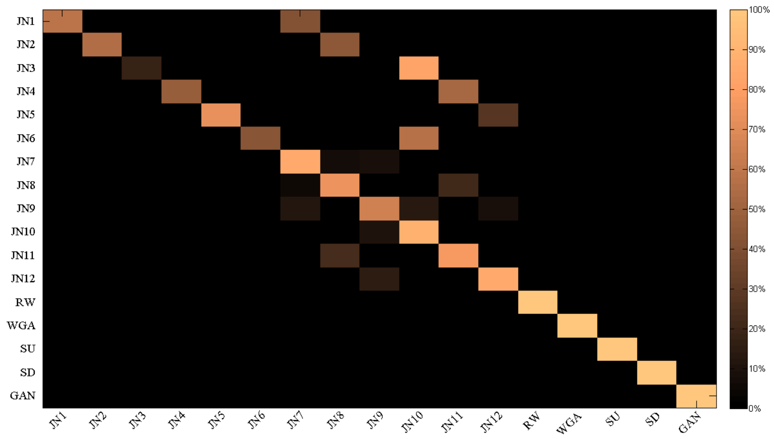


Figure 3.9: Confusion matrix for APs inferred by HHMM Model at level 2 using synthetic data

using the support of the walker, then getting out of bedroom and walking towards the living room (**LIV**). Once in the living room the user would sit down (**SD**) and tell the walker to go away (**WGA**). The APs involved during the navigational ADL would then be to traverse towards **jn 7**, followed by **jn 8** and **jn 11**, before the user reaches the living room. Figure 3.7 depicts synthetic data (in blue) of the four physical sensors (i.e. infra-red sensors and strain gauges) used to perceive basic human behaviours.

The ADLs were inferred at Level 1 with an accuracy of 90.74% by the HHMM framework, whereas the same was inferred with an accuracy of 54.03% by a L-DBN framework. Similarly, APs were inferred at Level 2 with an accuracy of 78.85% and 59.48% by the

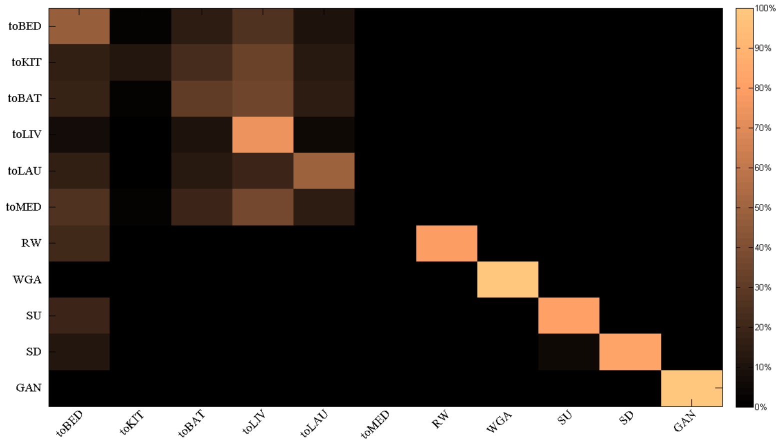


Figure 3.10: Confusion matrix for ADLs inferred by Layered DBN Model at level 1 using synthetic data

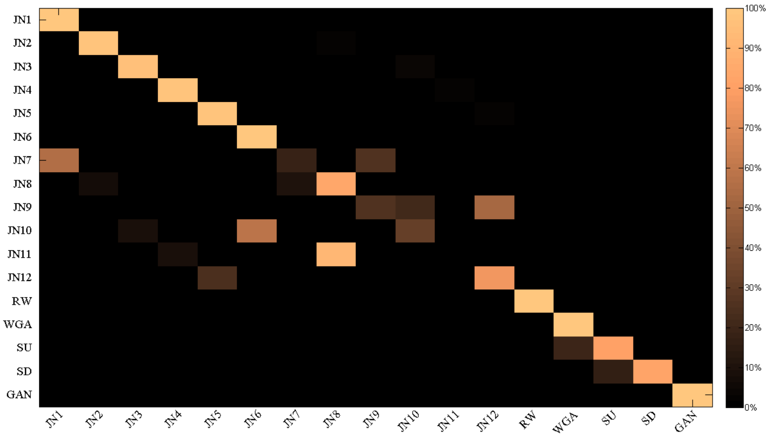


Figure 3.11: Confusion matrix for APs inferred by Layered DBN Model at level 2 using synthetic data

HHMM framework and the L-DBN framework respectively. The confusion matrix of the ADLs and APs inferred by both HHMM and L-DBN framework is depicted in Figure 3.8, 3.9 3.10, 3.11 respectively.

3.7.2 Evaluation with Data Collected using Power Walker

To further evaluate the framework with real-time data, we collected data from three healthy subjects while performing different ADLs using the power walker device depicted in Figure 3.5. The participants (one male, two female, (25-30 years of age)) did not have any technical background. They were also briefed and were given time to practice using

User Task	Start	End	APs (sub-activities)	Type of ADL
Going to Kitchen	Bedroom (BED)	Kitchen (KIT)	1 - 7 - 8 - 2	Navigational
Going to Bathroom	Kitchen (KIT)	Bathroom (BAT)	2 - 8 - 7 - 9 - 10 - 3	Navigational
Going to Bedroom	Bathroom (BED)	Bedroom (BED)	3 - 10 - 9 - 7 - 1	Navigational
Going to Living Room	Bedroom (BED)	Living Room (LRO)	1 - 7 - 8 - 11 - 4	Navigational
Going to Kitchen	Living Room (LRO)	Kitchen (KIT)	4 - 11 - 8 - 2	Navigational
Going to Laundry	Kitchen (KIT)	Laundry Facility (LAU)	8 - 7 - 9 - 12 - 5	Navigational
Going to Living Room	Laundry Facility (LAU)	Living Room (LRO)	5 - 12 - 9 - 7 - 8 - 11 - 4	Navigational
Stand Up (SU)	any location	same as start	Stand Up	Support activity
Sit Down (SD)	any location	same as start	Sit Down	Support activity
Recall Walker (RW)	Park location of walker	where it last left the user	Recall Walker	Navigate locally in a room (support)
Walker go Away (WGA)	Location where user sits down	Park location of walker	Walker Go Away	Navigate locally in a room (support)
General Assistive Navigation (GAN)	locally in a room	locally in a room	General Assistive Navigation	Assist user to navigate locally in a room (support)

Table 3.1: List of ADLs performed by a typical walker user. Paths represents the topological junction points visited to reach the goal destination

the walker so as to understand its functionality. We collected data for some of the many ADLs listed in Table 3.1 that a typical walker user would encounter. Figure 3.6 shows the trajectory plot of the path travelled by one of the users to visit various locations of interest to perform different ADLs. The data were manually labelled for cross validation and divided into two equal sets for training and testing purpose.

The ADLs were inferred at Level 1 with an accuracy in the range of 98% by the HHMM framework, whereas the same was inferred with an accuracy of 61% by the L-DBN framework(comparison shown in Table 3.2). The sequence of APs was inferred with an accuracy of 81% and 67% by the HHMM framework and L-DBN framework respectively. The confusion matrix of ADLs and corresponding APs inferred by both HHMM and L-DBN model is depicted in Figure 3.12, 3.13, 3.14 and 3.15 respectively.

3.7.3 Validating Synthetic Data with Real Walker Data

To validate the generated synthetic data with that of real walker data, we conducted an experiment where the HHMM framework was trained with synthetic data and tested with real walker data. The HHMM framework inferred ADLs with an overall accuracy of 98% whereas the APs were inferred with an accuracy of 75%. The test was performed

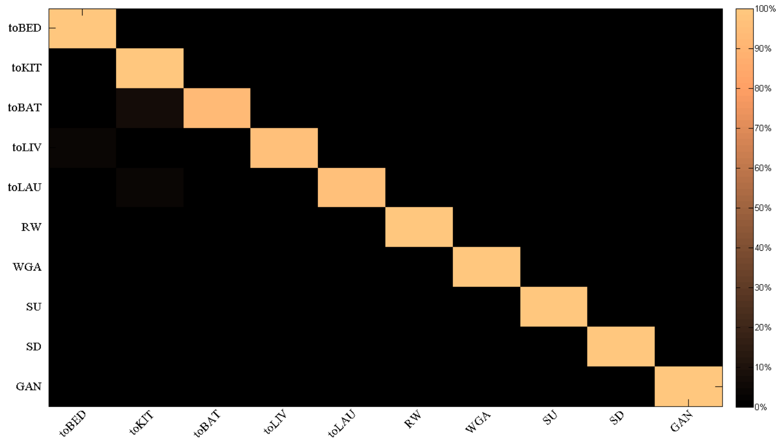


Figure 3.12: Confusion matrix for ADLs inferred by HHMM Model at level 1 using real walker data

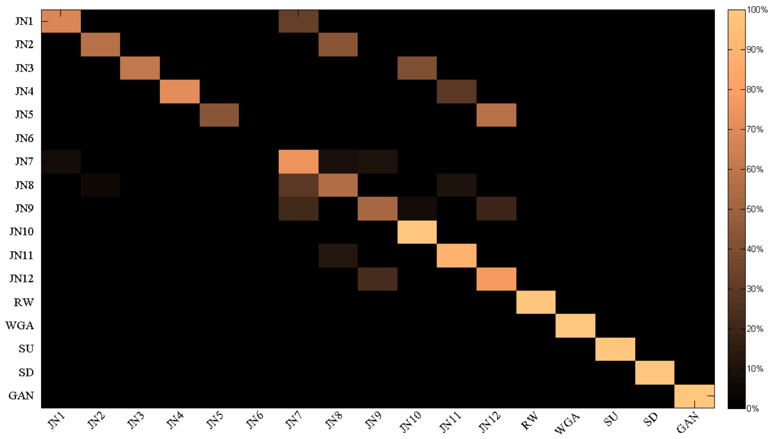


Figure 3.13: Confusion matrix for APs inferred by HHMM Model at level 2 using real walker data

only with the HHMM framework as the previous results in Sections 3.7.1 and 3.7.2 suggested that the L-DBN framework was unable to infer both ADLs and APs with high accuracy. The inference accuracy for testing the model with real data is higher than the same when tested with simulated data; this is due to the fact that the number of places visited using real data (7 locations) was less than those using synthetic data (30 locations).

Training the model with synthetic data and testing it with real walker data resulted in high inference accuracy, substantiating the hypothesis that the synthetic data is representative of the real data. Further, our evaluation of the proposed hierarchical

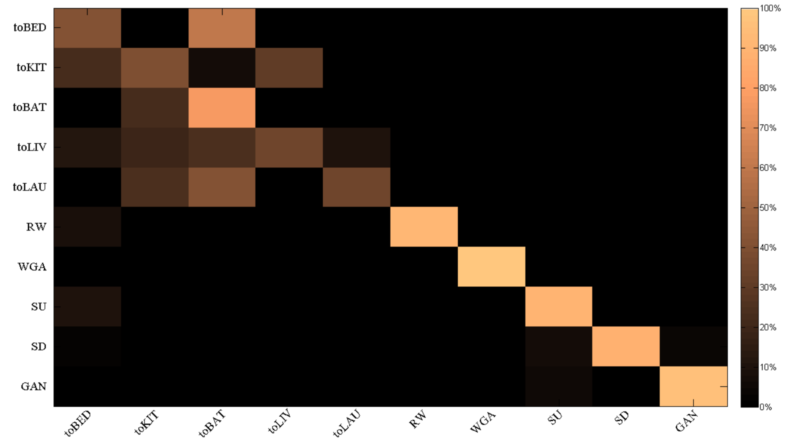


Figure 3.14: Confusion matrix for ADLs inferred by Layered DBN Model at level 1 using real walker data

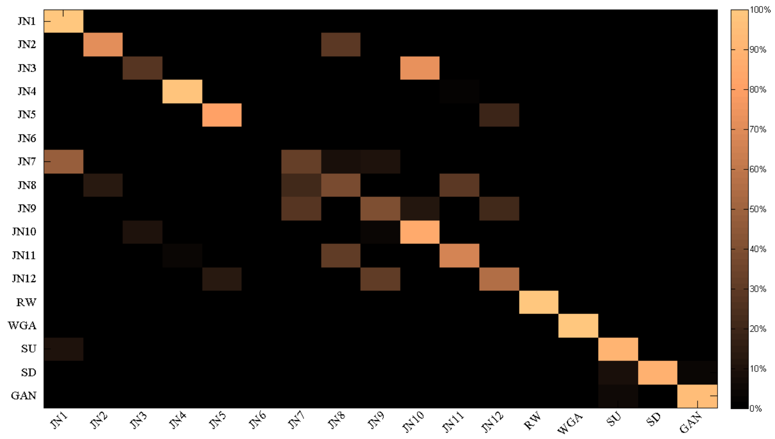


Figure 3.15: Confusion matrix for APs inferred by Layered DBN Model at level 2 using real walker data

framework seems to confirm the robustness of the model which infers the learned ADLs with very high certainty. A comparison of the synthetic data against real sensor data of four physical sensors is shown in Figure 3.7.

3.7.4 Comparison with Discriminative models

We also compared the accuracy of the HHMM and L-DBN frameworks with a hybrid HHMM/SVM model. Support Vector Machines (SVM) are class of powerful algorithms derived from statistical learning theory and applicable to pattern recognition problems. SVM efficiently constructs and trains the optimal separating hyper-planes in the

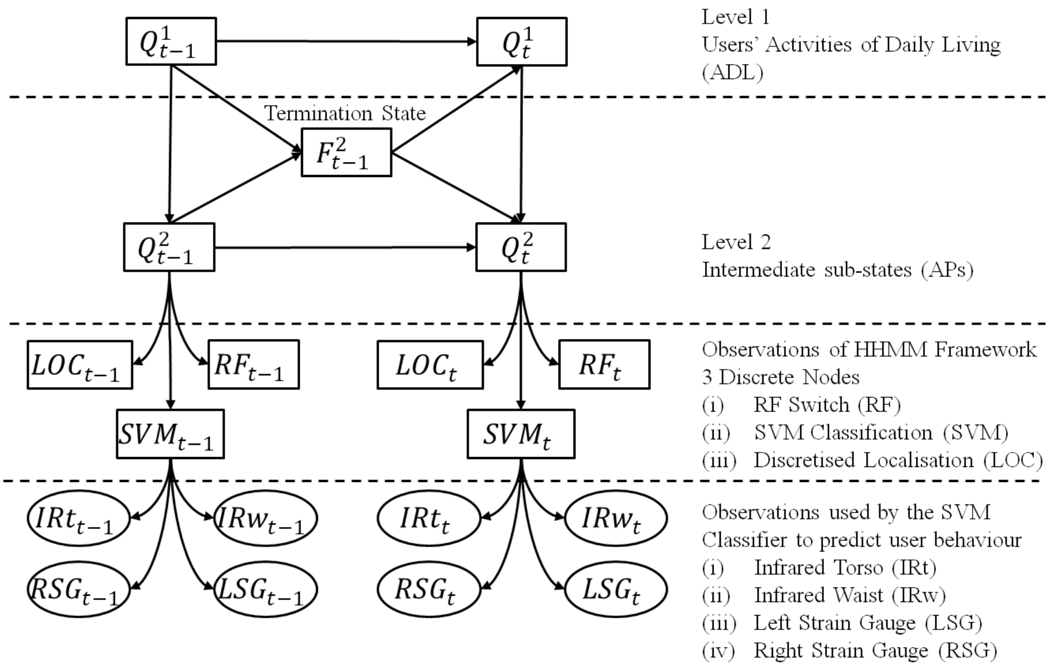


Figure 3.16: HHMM/SVM Hybrid model used to infer ADLs. The SVM classifies static user behaviour which acts as observations to the HHMM model which in turn are merged with other observation such as *LOC* and *RF* to infer the APs

	Activities of Daily Living (ADLs)										
Model/Activity	tobed	tokit	tobat	toliv	tolau	RW	WGA	SU	SD	GAN	Overall
L-DBN	40.55	40.23	76.99	34.74	35.22	91.51	99.70	90.10	87.50	95.13	60.95
HHMM	100	100	92.92	96.33	95.54	100	98.47	100.00	99.05	99.56	97.91
HHMM-SVM	100	100	100	100	100	100	99.39	99.03	98.10	99.85	99.83

Table 3.2: Inference accuracy of Generative and Discriminative Models (%)

kernel-induced feature space while enforcing the learning biases suggested by generalisation theory. Since SVM unlike HHMM and DBN, lack the ability to model time series researchers have combined them with HMM in various applications [Bishop and Lasserre, 2007] [Castellani et al., 2004] [Valstar and Pantic, 2007] [Stadermann and Rigoll, 2004], allowing the excellent discrimination performance of SVM to complement the temporal modelling properties of HMM and provide a higher inference accuracy. The HHMM/SVM hybrid model used in this work is shown in Figure 3.16. In the hybrid model the sensor information is used by the SVM to classify whether the user activity belongs to a navigational ADL or to one of the support activities. The SVM classification is merged with other sensor information (*RF* and *LOC*) which is then

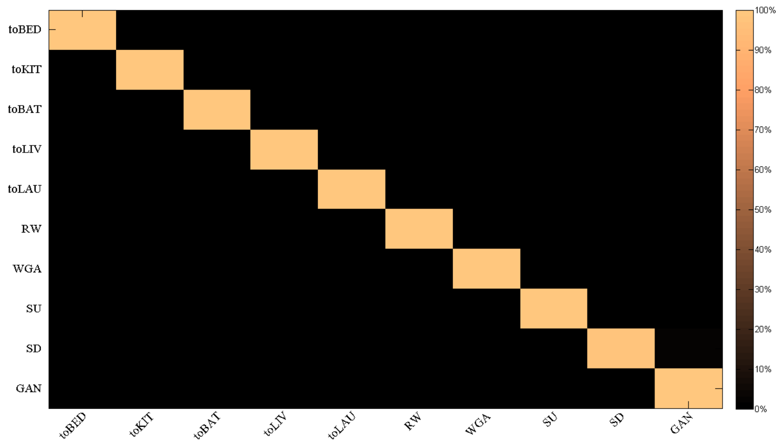


Figure 3.17: Confusion matrix for ADLs inferred by HHMM/SVM Hybrid Model at Level 1 using real walker data

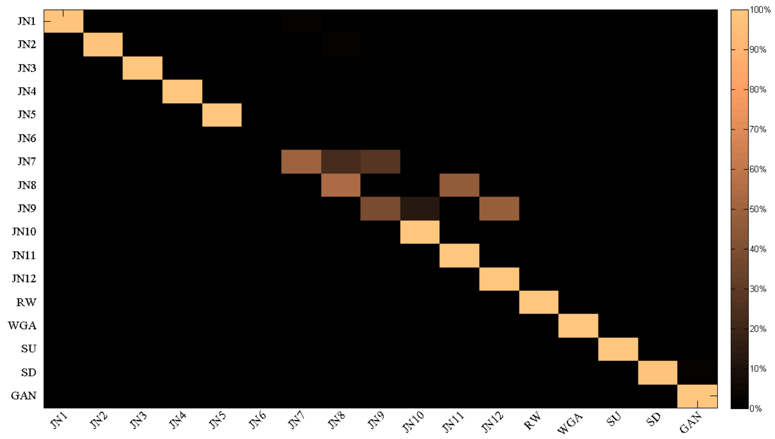


Figure 3.18: Confusion matrix for APs inferred by HHMM/SVM Hybrid Model at Level 2 using real walker data

used by the HHMM framework to exploit the temporal relations in the same fashion as before to infer APs and ADLs. It should be noted that the sensor data used for the SVM classifier at each time step was the same as those used by the HHMM and L-DBN framework.

The inference accuracy of the HHMM/SVM hybrid model to infer ADLs and the corresponding APs was around the same values as the proposed HHMM framework (approximately 99% and 81% respectively). The confusion matrix for HHMM/SVM hybrid model to infer ADLs and APs is shown in Figure 3.17 and 3.18 respectively. This further suggests that the proposed hierarchical structure is able to take full advantage

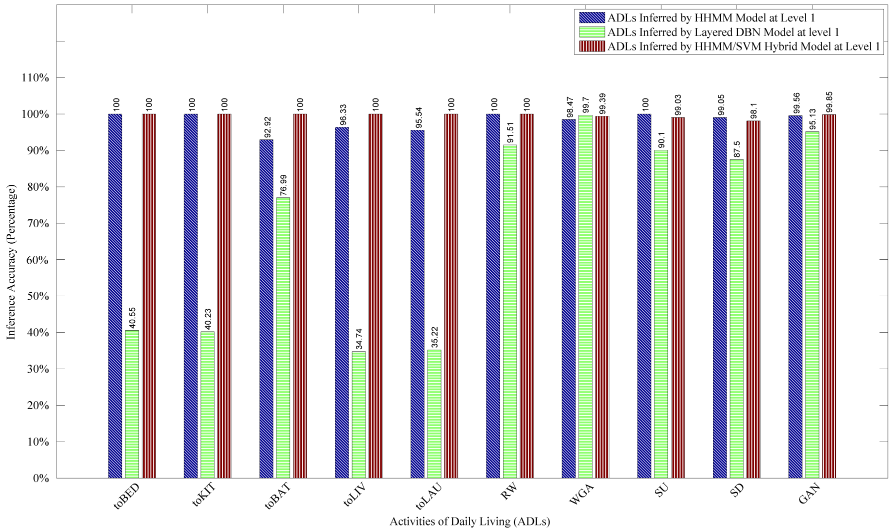


Figure 3.19: Comparison of ADLs inferred by HHMM Model, HHMM/SVM Model, Layered DBN Model and SVM classifier

of the temporal nature of ADLs. Yet by employing HHMM there is the significant advantage of not having to resort to (partial) supervised learning. This is a significant advantage for the target audience, as the physical abilities of the intended users will deteriorate with age. Additionally the network itself benefits from the capacity to adapt to ADLs changes without the need for expensive data tagging. A comparison of the inference accuracy of all three models is shown in Figure 3.19 and Table 3.2.

Further, examining the results in Table 3.2 and Figure 3.19, it can be seen that all the models inferred or classified the support activity (e.g. **RW, WGA, SU, SD, GAN**), with a very high accuracy where as the navigational ADLs were inferred with higher accuracy by the HHMM and HHMM/SVM hybrid model.

3.7.5 Inferring ADLs and APs with HHMM framework using On-line Inference

The inference results discussed in the previous Sections was performed off-line, i.e. the entire ADL sequence ($t = 1 : T$) was used to infer the APs and the respective ADL at any given time t ($t < T$). Such inference model can be useful for applications whereby walker users are asked to perform fixed routine ADLs, e.g. as part of their everyday routine

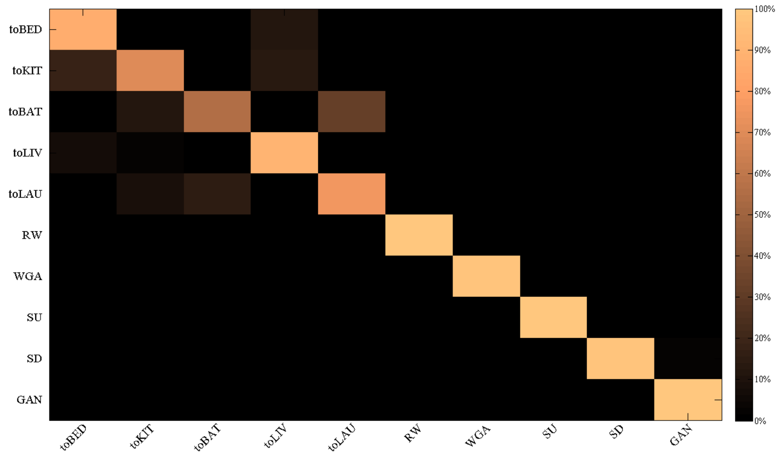


Figure 3.20: Confusion matrix for ADLs inferred using on-line inference by HHMM Model at level 1 using real walker data

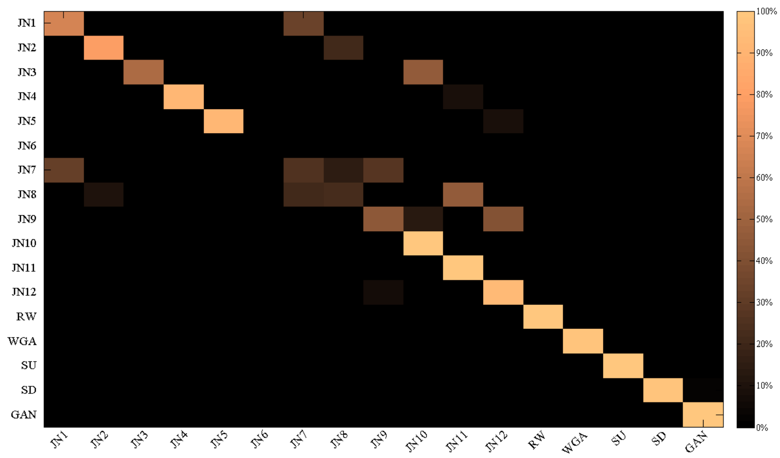


Figure 3.21: Confusion matrix for APs inferred using on-line inference by HHMM Model at level 2 using real walker data

exercises generally prescribed by the occupational therapist to maintain a healthier lifestyle or to follow a specific mobility rehabilitation program.

For the scenario of inferring the ADLs and the corresponding APs in real time (i.e. inferring the user ADLs based on the observation sequence till the current time) to provide assistance to the user as and when required, we utilised the forward algorithm described in Chapter 2. Note that the inference was performed using only the HHMM framework as the L-DBN model was shown to infer ADLs with lower accuracy in the previous experiment. The HHMM framework inferred ADLs with 87% accuracy, whereas the corresponding APs were inferred with 72%. The confusion matrix for ADLs and

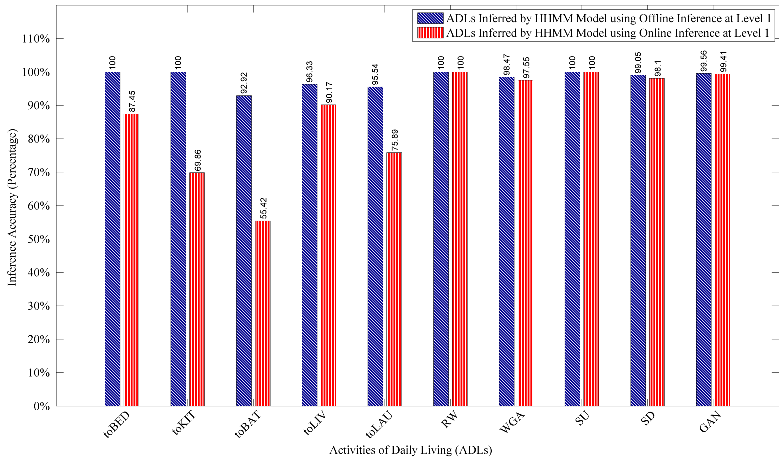


Figure 3.22: Comparison of ADLs inferred by HHMM Model using off-line inference and on-line inference technique

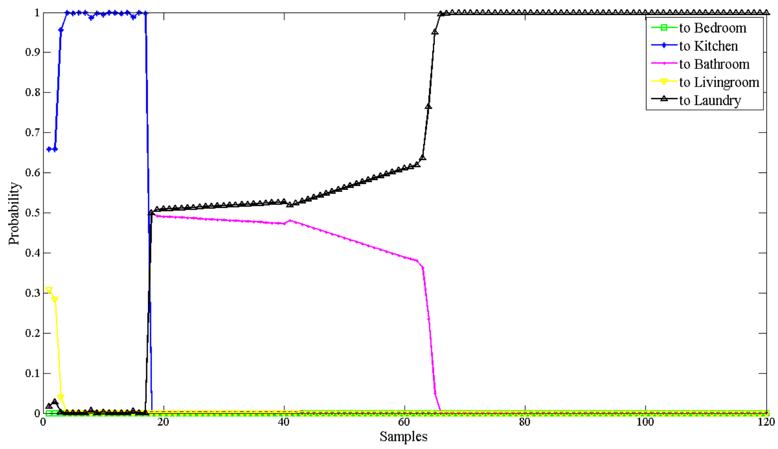


Figure 3.23: Probability inference of performing the ADL of going to the Laundry from Kitchen. Note that the inference is performed using the observation available till the current time.

APs are depicted in Figure 3.20 and 3.21 respectively. Furthermore, a comparison of the ADLs inferred by the HHMM framework using the off-line approach as compared to on-line approach is shown in Figure 3.22. The inference accuracy reduces by about 11% using the on-line approach as the observation information used is limited to that available till the current time. As an example, the inference probability evolution for the activity of going to the living room from the laundry is shown in Figure 3.23.

3.8 Summary

In this chapter we presented a solution to model ADLs encountered by a typical mobility aid device user in everyday life. The work is an extension of the initial results published in [Patel et al., 2012]. Our proposed approach of modelling ADLs by decomposing them into APs was evaluated with both synthetic and real data collected from users of an automated power walker. Using synthetic data, the ADLs were inferred with an accuracy of 90% using the HHMM framework, while an overall accuracy of 98% was obtained using real data collected using the power walker. This high level of accuracy attained by using the HHMM framework was due to its capability of representing ADLs at multiple levels.

To further support our proposition of using a hierarchical framework to model ADLs, we compared the inference accuracy of HHMM model with other machine learning techniques, such as a layered DBN model and a hybrid HHMM/SVM model. The results demonstrate that typical support activities are inferred with similar accuracies by all models. However, prediction accuracy for navigational activities which are long term in nature is almost halved for the L-DBN as compared to HHMM models, further proof that decomposition of ADLs at multiple levels and the inherent temporal information present in ADLs plays a critical role in predicting these activities. On the other hand, the HHMM/SVM hybrid model inferred ADLs with the same accuracy compared to the HHMM model. This is because the SVM classification of APs at the intermediate level is further utilised by the HHMM model to exploit the temporal and hierarchical relationship between APs and ADLs. Lastly, we also compared the accuracy of inferring ADLs using off-line and on-line inference algorithm for the HHMM framework.

The work presented in this chapter assumes that a topological representation of the given environment is available. The junction points in the topological map act as APs which were combined in different sequence to define the activities of visiting location of interest. Generating a topological representation of a given environment becomes a cumbersome process as each environment will have its own unique topological map. In the next chapter we present a novel technique, where we model ADLs such that the use of a topological map can be made redundant.

Chapter 4

Modelling Activities of Daily Living using Human Motion Models

4.1 Introduction

In the previous chapter we demonstrated a Hierarchical Hidden Markov Model (HHMM) based probabilistic framework to model a broad spectrum of Activities of Daily Living (ADLs) as performed by a mobility aid device (walker) user. However, the approach used in the previous chapter assumes that the topological representation of a given environment is available where locations of interest are connected using junction points and edges. This assumption requires the additional work of generating a topological representation of the user environment *(private home, old age home or hospitals).*

In this chapter, we present a motion primitive based modelling technique that captures the local human behaviour which represents the person's interaction with the robotic device and the environment, while performing ADLs. The model follows a similar technique as employed in Chapter 3, which consists of decomposing complex ADLs into a string of Action Primitives (APs). The primary advantage of using motion primitive

(a) Power Walker (b) Robotic Wheelchair

Figure 4.1: (a) Power Walker and (b) Robotic Wheelchair used to model ADLs

based modelling is its ability to model ADLs and the corresponding APs by using human motion and therefore does not require any topological representation of the environment. We evaluated the proposed framework with data collected with two different mobility devices: a power walker and a robotic wheelchair (Figure 4.1). We also compared the inference accuracy of the HHMM Model with that of a Layered Dynamic Bayesian Network (L-DBN) and a 2 stage SVM classifier. The L-DBN and a 2-stage SVM classifier were utilised to allow for a fair comparison with the HHMM Model. In all the three frameworks, ADLs are modelled at the higher level and inferred from APs and the APs are in-turn modelled at the intermediate level and are inferred using the physical sensors.

4.2 Related Work

The approach of motion primitive based activity recognition has become an important focus in research due to its ability to capture the local characteristics of activity signals [Zhang and Sawchuk, 2012]. The motion primitive based models are inspired by human speech signals [Ghasemzadeh et al., 2008]. In human speech, sentences are first divided into isolated words, which are then divided into a sequence of phonemes. Following the same idea, in motion primitive based model, each activity is represented as a sequence of APs which acts as the smallest unit to be modelled. Stiefmeier and colleagues constructed motion primitives by dividing the activity trajectory into fixed-length windows with an identical spatial duration, where each window was mapped to a motion primitive based on its trajectory direction in the Cartesian space [Stiefmeier et al., 2007]. Nguyen and colleagues used a HHMM framework to model and recognise three human activities (e.g. having a meal) in a confined space, while tracking the user with two static cameras [Nguyen et al., 2005]. The semantics embedded in the activities were tightly coupled to the locations where the relevant objects of interest - such as fridge, cupboard etc. - were located, not the actual interactions between user and objects. Osentoski *et. al.* used an Abstract Hidden Markov Model (AHMM) to models behaviours in an indoor environment [Osentoski et al., 2004]. The proposed model decomposed user states into intermediate states which consisted of small clusters formed by dividing the entire action trajectory. In the context of an outdoor environment Liao and colleagues proposed a surveillance system using GPS sensors to infer user's daily activities in a large and complex environment [Liao et al., 2007]. They used a HHMM framework to infer user's mode of transportation and the destination location, and to predict both short and long term movements. The framework was also able to infer if the user was deviating from normal activities as an indicator to provide guidance cues.

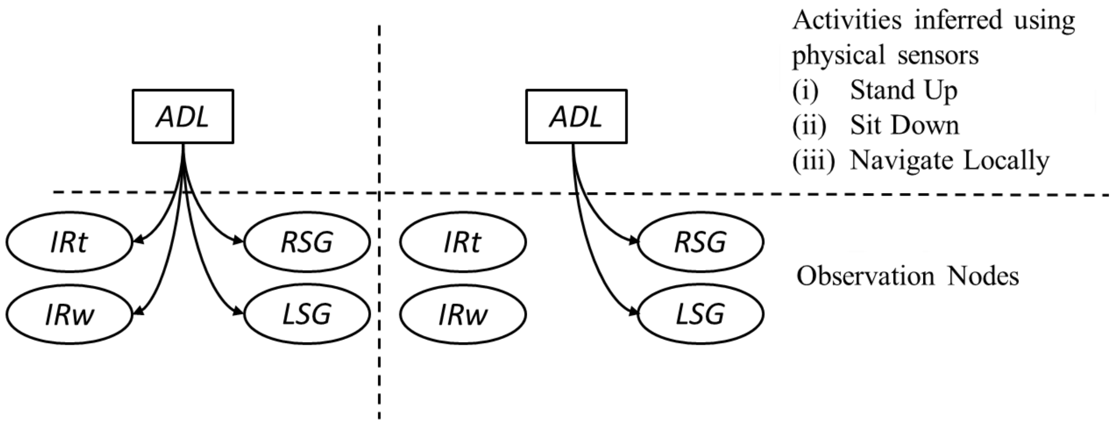
In the context of mobility aid device users, researchers have developed different frameworks capable of modelling ADLs based on basic human behaviour. Alwan *et. al.* describe a method that assesses basic walker-assisted gait characteristics, including heel strikes, toe-off events, stride time, double support and right and left single support phases [Alwan et al., 2005]. These statistics were based on the measurements of weight transfer between the user and the walker as perceived by the sensor (in the form of two load cells fitted on the handles of the walker). A simple threshold approach was utilised to detect peaks and valleys in the load measurements, which are assumed to be indicative of certain events in the gait cycle. This work mainly focusses on low level gait statistics. Hirata et al. [Hirata et al., 2006] provided a framework which would estimate user states. They recognise three user states: walking, stopping and emergency (including falling). These states are inferred based on the distance between the user and the walker and the velocity of the walker.

Similar work has been done using a power wheelchair as a mobility device. Wheelchairs have been prescribed to people who do not have upper body strength and are unable to maintain gait balance while walking. Fehr *et. al.* reported that 40% of wheelchair users find steering nearly impossible [Fehr et al., 2000]. This is due to the fact that constant vigilance is required on the part of the user to sense their environment, recognise hazards, and be able to transfer their desired motion into continuous joystick commands for the wheelchair [Brose et al., 2010]. Researchers [Pruski et al., 2002] [Demeester et al., 2008] [Carlson and Demiris, 2010] [Atrash and Pineau, 2009] have developed a number of strategies whereby the user's intention to perform an activity is perceived through a different sensor system which is further utilised by the intelligent model to reduce the overall load on the user to operate the wheelchair. Prunski and colleagues proposed a control strategy for users of an intelligent wheelchair that provides control assistance that best suits the user behaviour. The control system provides the necessary assistance to control the motion of the wheelchair and align it to the activity the user wants to perform. Carlson and Demiris proposed a collaborative control system that would provide assistance as and when required thereby reducing the concentration required by the user as well as their overall workload. Examples of such scenarios would be passing through a narrow doorway, manoeuvring in cluttered environment, or in situations when the user is engaged in some secondary activity like talking to someone. The proposed collaborative control model was able to provide adaptive assistance and additional safety from a dynamic obstacle avoidance module [Carlson and Demiris, 2010]. Similarly, Vanhooydonck *et. al.* [Demeester et al., 2008] proposed a shared control strategy where the intelligent algorithm provides low level assistance from high level user intentions as inferred by the intelligent algorithm.

In this chapter, we propose to model complex ADLs of two mobility aid devices, a power walker and a robotic wheelchair using basic human motion as perceived by the physical sensors fitted on the mobility devices. This was done by exploiting both the temporal relationship between *ADLs* and *APs*, and the spatial relationship between *APs* and *observation* using a probabilistic framework. The dictionary of APs was developed based on basic human motion. For example, in the case of modelling navigational ADLs, the human motion related APs will consists of the person's navigational intention of going in a specific direction, *turning left or right*. The probabilistic framework in the form of HHMM is utilised to learn and assist with an array of the ADLs that a typical user of locomotive supportive devices would normally engage in. The type of interactive ADLs considered include for instance 'standing up', 'going to the kitchen', 'recalling the platform', etc. The complex ADLs are further decomposed into a string of meaningful APs which are based on human motion. For instance the task of 'going to the kitchen' can be decomposed into directional APs of going straight in a specific direction followed by turning right or left depending on the route to be followed in a given environment.

4.3 Mobility Device to Support Activities of Daily Living

We focus on modelling ADLs performed by users of two common mobility devices: a power walker and a robotic wheelchair (Figure 4.1). Both these devices are prescribed to people having poor upper body strength, poor gait stability as well as other factors. The scope of ADLs performed by users of both these devices primarily extends to manoeuvring at different locations of interest (e.g. kitchen, bedroom, bathroom, etc.) and performing other support activities such as standing up using the support of the mobility device (for power walker user). The ADLs are modelled using human behaviour as perceived through different physical sensors installed on the mobility device.



(a) Manually Defined Graphical Model (b) Graphical model evolved using Structure Learning

Figure 4.2: (a) Graphical Model defined manually to infer local human behaviour/activities (b) Graphical Model evolved using bayesian structure learning approach to infer local human behaviour/activities

4.3.1 Power Walker as a Mobility Device

The power walker (Figure 4.1(a)) used to model ADLs is the same as that used for experiments in Chapter 3. Further hardware and software details regarding the platform can be found in Appendix A and C respectively.

Sensor Dependencies to Model basic Human Behaviour

The physical sensors, two infra-red (IRs) proximity sensors and the strain gauges (SGs) installed on the walker are mainly used to monitor the basic human behaviours such as *sitting down, standing up, and navigation*, which are then fused with other sensor information to model complex ADLs. Through visual inspection of the data, we observed that the infra-red sensor data did not provide any extra information to model local human behaviours. To further validate our belief we used Structure Learning (SL) to exploit the dependencies between physical sensors and the state of human behaviour. We utilised a Bayesian Network (BN) structure learning tool kit called WinMine [Chickering, 2002] developed by Microsoft Research to learn the structure and parameter dependencies of the model. WinMine uses a greedy directed acyclic graph based algorithm starting with the model containing no edges which are then greedily added, deleted and reversed based on the parameters, until a local maximum is reached. The evolved structure is then evaluated by calculating the log posterior probability for each of the hidden/output nodes. The structure that evolved after performing structure learning using WinMine toolkit is shown in Figure 4.2(b). To further validate sensor dependencies to infer user behaviour, we compared the inference accuracy of the graphical model evolved using structure learning (Figure 4.2(b)) to the accuracy of the manually defined graphical model (Figure 4.2(a)). Both models inferred the user behaviour with the same accuracy of 99%. Hence in this chapter we model ADLs using the sensor information from the Left and Right Strain Gauges (*LSG, RSG*) only. The information available from these two physical sensors was sufficient to perceive human behaviour, which is further combined with other observations to model the ADLs performed by a typical walker user.

4.3.2 Robotic Wheelchair as a Mobility Device

The wheelchair used for experimentation (Figure 4.1(b)) is a commercially available power wheelchair (Invacare rollar M1 [Invacare, reviewd on 3rd January 2013]) modified with necessary hardware such as wheel encoders, a Hokuyo URG-04LX laser range finder, RF switch and a digital-to-analog (DAC) interface unit. Further details of hardware and software integration of the wheelchair can be found in Appendix B and C respectively.

It should be noted that the assistance provided by a wheelchair used in this work is mainly navigational in nature. The user's behaviour when performing a specific ADL is perceived through the joystick of the wheelchair, where the user provides navigational cues of where/which direction the user intends to go. The navigational cues provided by the user are fused with other information such as localisation and RF switch to model APs, which are then combined in different sequence to model the overall ADL the user is trying to perform.

Action Primitive	Abbrev.	Description
Recall Walker/Wheelchair	RW	Recalling walker/wheelchair to use it
Walker/Wheelchair Go Away	WGA	Instruction the walker/wheelchair to go way so that it is not an hindrance
Stand Up	SU	Stand up using the support of the walker (only for walker)
Sit Down	SD	Sit down using the support of the walker (only for walker)
Go Straight Northwards	GSN	Going straight with orientation towards north
Go Straight Westwards	GSW	Going straight with orientation towards west
Go Straight Eastwards	GSE	Going straight with orientation towards east
Go Straight Southwards	GSS	Going straight with orientation towards south
Turn Left	TL	Turning left
Turn Right	TR	Turning right

Table 4.1: Action Primitives to perform various ADLs

4.3.3 Modelling ADLs of Mobility Device User

The ADLs modelled for users of different support devices are in the context of a typical home environment where a walker or wheelchair user would generally have a well-known set of locations that they visit during their daily activities, e.g. kitchen, bedroom, bathroom, laundry, etc. or activities performed at a single location, e.g. standing up from sitting position (confined to walker platform). ADLs can broadly be categorised into two group: firstly, support activities which are static in nature and are generally associated with being performed at a single location (*e.g. stand up, sit down, recall mobility device*) and secondly, navigational ADLs which involve visiting location of interest (*e.g. going to kitchen*). Support activities being in their most primitive state already cannot be decomposed further, hence such ADLs are also regarded as independent APs in themselves. On the other hand, navigational activities can be further decomposed into more basic navigational components *(e.g. turn left)*. A typical scenario could be illustrated by considering for instance the activities involved in going to the kitchen from the bedroom (shown in light blue in Figure 4.6). The sequence of activities can be decomposed as "*Recall Walker/Wheelchair* ⇒ *Stand Up (walker)/transfer to wheelchair* ⇒ *Navigate in Room* (until out onto the corridor) ⇒ *To Kitchen* [Go Straight Southwards ⇒ Turn Right ⇒ Go Straight Westwards (until entering kitchen)] ⇒ *Navigate in Room* ⇒ *Sit Down (walker)/ transfer from wheelchair* ⇒ *Walker/Wheelchair Go Away*" (The

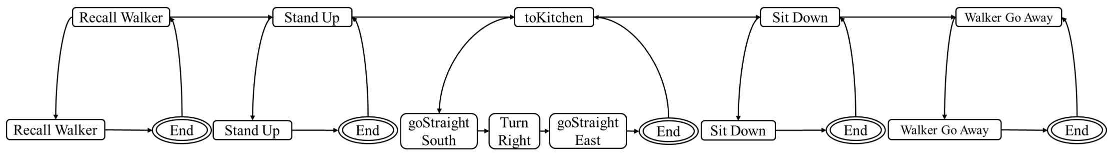


Figure 4.3: Time series of activities performed in sequence and their further decomposition into sequence of APs

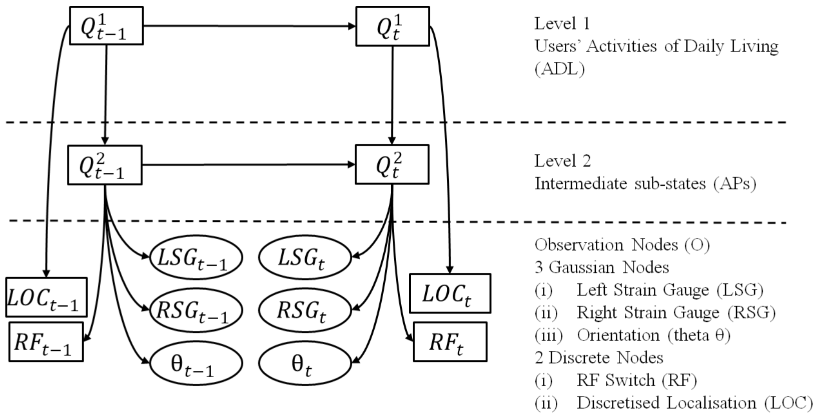
decomposition of the activity into navigational and support ADLs and their respective transitions from one ADL to another is shown in Figure 4.3)

4.4 Probabilistic Models to Predict ADLs using Human Motion Models

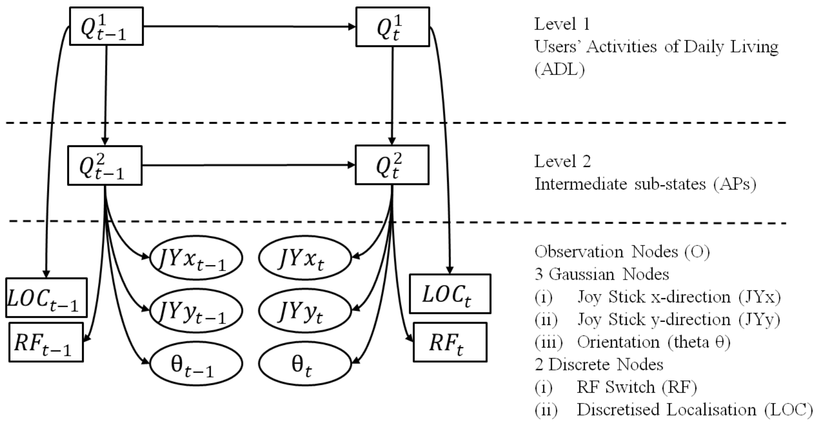
We employed a HHMM framework to model ADLs using APs which are based on the basic human motion. The HHMM model exploits the decomposition of ADLs into a string of APs which consists of pool of intrinsic human behaviour. A comparison of the inference accuracy of the HHMM model was also done with a Layered Dynamic Bayesian Network (L-DBN) and a 2-stage SVM classifier. All the three models i.e. HHMM, L-DBN and the layered SVM classifier, infers APs at the intermediate level using the sensor observation whereas the ADLs are inferred at the higher level by combining the APs in different sequences.

4.4.1 Layered Dynamic Bayesian Network (L-DBN)

A 2 layer DBN framework which models ADLs and the corresponding APs for both the mobility devices is shown in Figure 4.4. The ADLs and APs correspond to the hidden state of the L-DBN structure, whereas the sensor readings corresponds to the observed nodes. The APs are inferred (at Level 2) using observation whereas the ADLs are inferred (at Level 1) using both the inferred APs at level 2 and observation from the sensors. The hidden nodes are discrete whereas the observed nodes consists of both discrete and continuous nodes. The two slice L-DBN model is unrolled infinitely. The only connection between each time slice is via the hidden states at Level 1 and Level 2.



(a) Layered DBN Model used to infer ADL for Walker user



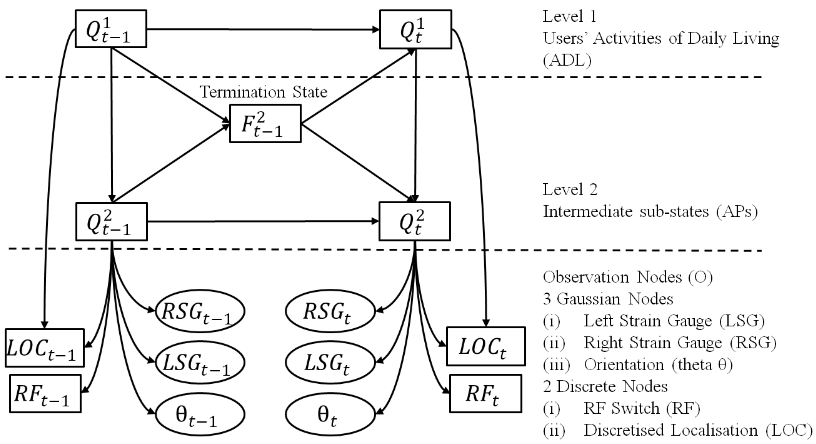
(b) Layered DBN Model used to infer ADL for Wheelchair user

Figure 4.4: Layered DBN Model used to infer ADL for walker user (Figure (a)) and wheelchair user (Figure (b))

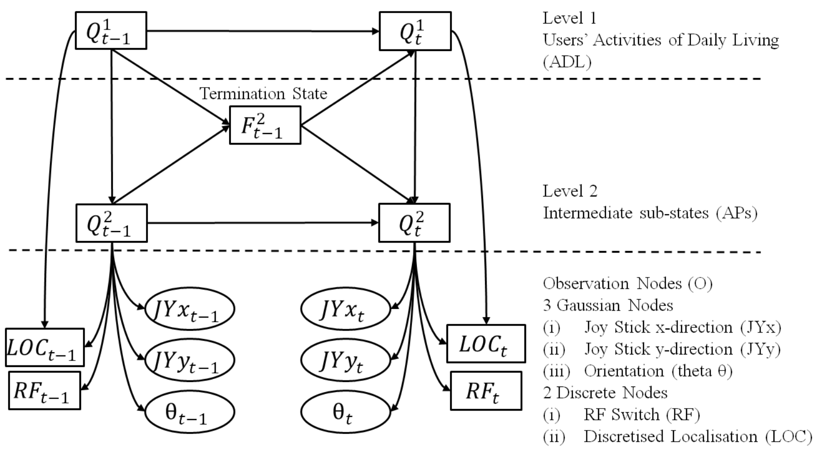
Probabilities in the form of prior probabilities ($P(Q^1_1)$ & $P(Q^2_1)$), transition probabilities ($P(Q^1_t|Q^1_{(t-1)})$ & $P(Q^2_t|Q^2_{(t-1)}, Q^1_t)$) and observation probabilities ($P(O_t|Q^1_t, Q^2_t)$) are provided at both the levels, which are further optimised by learning the model parameters from the data.

4.4.2 Hierarchical Hidden Markov Model (HHMM)

The HHMM framework used in this experiment for both the mobility platforms is depicted in Figure 4.5. The user activities are hierarchically split-up at different levels of abstraction. Similar to the L-DBN model, the ADLs are inferred at the top level (Level 1) whereas the APs are predicted at the intermediate level (Level 2). Along with



(a) HHMM Model used to infer ADL for Walker user



(b) HHMM Model used to infer ADL for Wheelchair user

Figure 4.5: HHMM Model used to infer ADL for walker user (Figure (a)) and wheelchair user (Figure (b))

the graphical model the prior, transition, observation and termination probabilities are defined for the HHMM model which is further optimised using the Expectation-Maximisation (EM) learning algorithm. The prior probabilities at both the levels are defined using Equation 4.1, the transition probabilities at Level 1 and 2 are defined using by Equation 4.2 and Equation 4.3 respectively, while the termination probabilities are defined using Equation 4.4.

$$
\begin{aligned}
P(Q^1_1) &= \pi^1(j) \\
P(Q^2_1) &= \pi^2_k(j)
\end{aligned}
\tag{4.1}
$$

$$P(Q_t^1 = j|Q_{t-1}^1 = i, F_{t-1}^2 = f) = \begin{cases} A^1(i,j) & \text{if } F_{t-1}^2 = 0 \\ \pi^1(j) & \text{if } F_{t-1}^2 = 1 \end{cases} \tag{4.2}$$

$$P(Q_t^2 = j|Q_{t-1}^2 = i, F_{t-1}^2 = f, Q_t^1 = k) = \begin{cases} A_k^2(i,j) & \text{if } F_{t-1}^2 = 0 \\ \pi_k^2(j) & \text{if } F_{t-1}^2 = 1 \end{cases} \tag{4.3}$$

$$P(F_t^2 = 1|Q_t^1 = k, Q_t^2 = i) = A_k^2(i, end) \tag{4.4}$$

In Equation 4.1, 4.2, A^1 and π^1 represent the transition and initial probabilities respectively at Level 1 where as in Equation 4.3 and 4.4 A_k^2 and π_k^2 represents the same at Level 2 given the state at Level 1 is k. The observation probabilities are defined as a mixture of Gaussian and/or discrete nodes. The probabilities of observation nodes which have dependencies at both the levels ((Q_1&Q_2) are defined by Equation 4.5 while Equation 4.6 defines probabilities of the node which have dependencies only on the intermediate node (Q_2).

$$\begin{aligned} P(O_t|Q_t^1 = i, Q_t^1 = j) &= N(\mu_{i,j}, \Sigma_{i,j}) \\ P(O_t|Q_t^1 = i, Q_t^2 = j) &= C(i,j) \end{aligned} \tag{4.5}$$

$$\begin{aligned} P(O_t|Q_t^2 = j) &= N(\mu_j, \Sigma_j) \\ P(O_t|Q_t^2 = j) &= C(j) \end{aligned} \tag{4.6}$$

In both the L-DBN and HHMM models the transition probabilities for interstate connectivity are defined based on common laws of operation as in what is perceived as accepted behaviour from the intended user pool (e.g. the user is unlikely to sit down immediately after standing up or it is impossible for the user to end up in state of go away after the user has been in state of going to the kitchen). These probabilities are conditioned further based on the training data used during the learning process.

The sensors used at the observation level consists of both physical sensors and the localisation information, details of which are as follows:

User Task	Abbrev.	Description
Recall Walker/Wheelchair	RW	Recalling walker/wheelchair to use it
Walker/Wheelchair Go Away	WGA	Instruction the walker/wheelchair to go way so that it is not an hindrance
Stand Up	SU	Stand up using the support of the walker (only for walker)
Sit Down	SD	Sit down using the support of the walker (only for walker)
to Bedroom	BED	Navigating to the bedroom
to Kitchen	KIT	Navigating to the kitchen
to Bathroom	BAT	Navigating to the bathroom
to Living room	LIV	Navigating to the living room
to Laundry	LAU	Navigating to the laundry
to Medical	MED	Navigating to the medical
Navigating in a room	GAN	Navigating locally in a room

Table 4.2: Users Activity of Daily Living (ADLs)

- 3 Gaussian nodes, two of which are the readings from the physical sensors installed on the walker (strain gauges *(LSG, RSG)*) and wheelchair (i.e. joystick *(JYx, JYy)*) and the orientation (θ) information available through the localisation software).
- 2 discrete nodes, RF and Localisation. A RF switch is used by the user to indicate the walker/wheelchair to go away and recall as needed. Location (metric (x, y)) is derived from a localiser and is discretised into 19 locations which are supplied as an observation to both L-DBN & HHMM framework.

4.5 Data Collection

Given the nature of the work, and the sensitivity of undertaking trials with elderly/frail subjects, healthy volunteers (25-31 years of age) were asked to participate in a set of experiments, and the data were collected for our initial validation purposes. The volunteers were briefed on the working of both the platforms and were also given time to practice using the platform so as to understand its functionality. During the experiments, the powered walker was configured to simply provide steering assistance based on data sensed from the handle strain gauges whereas the wheelchair was configured to

User Task	Start	End	APs (sub-activities)	Type of ADL
Going to Kitchen	Bedroom (BED)	Kitchen (KIT)	GSS - TR - GSW	Navigational
Going to Bathroom	Kitchen (KIT)	Bathroom (BAT)	GSE - TL - GSN	Navigational
Going to Bedroom	Bathroom (BAT)	Bedroom (BED)	GSS - TR - GSW - TL - GSN	Navigational
Going to Laundry Facility	Bedroom (BED)	Laundry Facility (LAU)	GSS - TL - GSE - TL - GSN - TR - GSE - TL - GSN	Navigational
Going to Medical Room	Bedroom (BED)	Medical Room (MED)	GSS - TL - GSE	Navigational
Going to Laundry Facility	Medical Room (MED)	Laundry Facility (LAU)	GS - TL - GSN - TR - GSE - TL - GSN	Navigational
Going to Medical Room	Bathroom (BAT)	Medical Room (MED)	GSS - TL - GSE	Navigational
Stand Up (SU)	any location	same as start	Stand Up	Support activity (for walker user)
Sit Down (SD)	any location	same as start	Sit Down	Support activity (for walker user)
Recall Walker/Wheelchair (RW)	Park location of walker/wheelchair	where it last left the user	Recall walker/wheelchair	Navigate locally in a room (support)
Walker/Wheelchair go Away (WGA)	Location where user sits down	Park location of walker/wheelchair	walker/wheelchair Go Away	Navigate locally in a room (support)
General Assistive Navigation (GAN)	locally in a room	locally in a room	General Assistive Navigation	Assist user to navigate locally in a room (support)

Table 4.3: List of ADLs performed by users of both the mobility devices. APs represents the navigational cues provided through the user's motion

Mobility Device	Model/Activity	RW	WGA	SU	SD	BED	KIT	BAT	LAU	MED	GAN	Overall
Walker	HHMM	100.00	90.63	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.57
	L-DBN	100.00	62.40	100.00	100.00	57.36	42.11	45.14	57.67	61.07	100.00	65.99
	SVM	100.00	99.22	100.00	100.00	55.96	0.00	88.81	92.97	61.07	98.10	79.93
Wheelchair	HHMM	99.72	83.82	N.A.	N.A.	100	98.84	100.00	100.00	100.00	99.50	98.54
	L-DBN	89.07	93.07	N.A.	N.A.	23.01	42.53	89.19	56.00	26.03	79.19	64.30
	SVM	99.72	90.20	N.A.	N.A.	54.42	45.09	71.88	82.58	5.12	99.33	75.06

Table 4.4: ADLs Inferred by Generative and Discriminative Models (in Percentage)

provide the same based on the data sensed from the joystick of the wheelchair.

In order to assist people in performing their everyday activities, it is important to understand the patterns a user might follow to accomplish a given activity. In the scope of this project, we pre-defined some of the many everyday activities a typical walker/wheelchair user would normally encounter, although under clinical tests these patterns may be better defined with the help of an occupational therapist [Practice, 2008], or extracted from actual raw data from the subjects when using any of the mobility devices in the home environments. Once defined, data were collected from the sensors while the user performed activities as listed in Table 4.2. The number of APs in which all these ADLs can be clustered is listed in Table 4.1. A complete list of ADLs performed by users of both platforms is shown in Table 4.3.

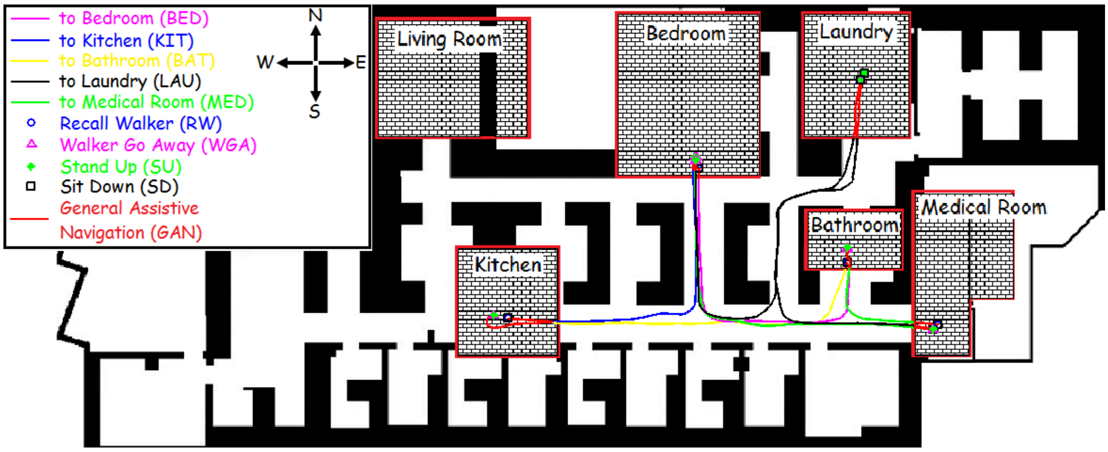


Figure 4.6: 2D Bird's eye view of environment divided into typical locations of interest in an home, superimposed with navigational routines performed by one of the walker user. Wheelchair data was also collected on the same navigational routines that a user would perform

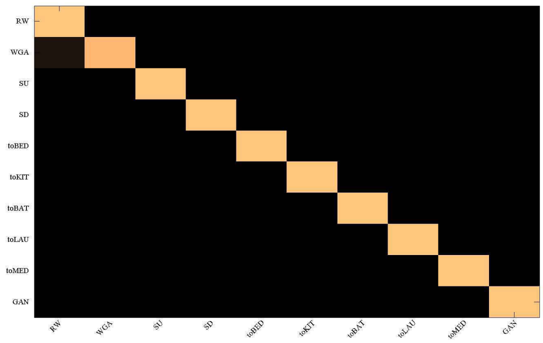
Mobility Device	Model/APs	RW	WGA	SU	SD	TL	TR	GSS	GSW	GSN	GSE	Overall
Walker	HHMM	100.00	90.63	100.00	100.00	83.91	51.38	94.76	93.16	91.06	86.03	88.88
	L-DBN	100.00	62.40	100.00	100.00	52.81	41.01	94.77	94.30	92.53	86.21	86.61
	SVM	100.00	99.22	100.00	100.00	90.80	84.53	99.32	100.00	99.04	98.90	98.09
Wheelchair	HHMM	99.72	83.82	NA	NA	96.01	97.50	98.24	81.37	97.56	89.48	93.73
	L-DBN	89.07	93.07	NA	NA	94.24	99.50	88.69	92.42	94.02	91.42	92.83
	SVM	99.72	90.20	NA	NA	97.10	98.25	98.82	99.24	97.83	98.71	97.91

Table 4.5: Inference accuracy of Action Primitives by Generative and Discriminative Models (in Percentage)

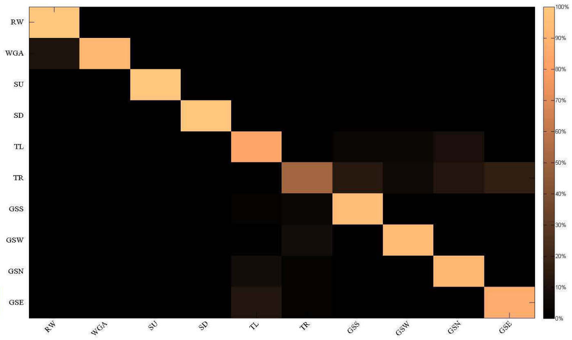
4.6 Results

The L-DBN and HHMM models were tested off-line using real time data collected from three users while performing various activities as described in Section 4.5 with both mobility platforms. User data was logged using Player/Stage open source toolbox [Gerkey et al., 2003]. Data were collected for different navigational routines and support activities as specified in Figure 4.6 to visit 5 location of interest.

We used the BNT toolbox [Murphy, 2002] to learn and infer user ADLs using both the probabilistic models. EM, an unsupervised mode of learning was used to learn user activities and Maximum Likelihood Estimator was used for inference. The data were manually labelled for cross validation and were divided in two equal sets for the purpose of training and testing. Using the HHMM framework, the inference accuracy for the ADLs for the walker and wheelchair was 99.57% and 98.54% respectively while it was

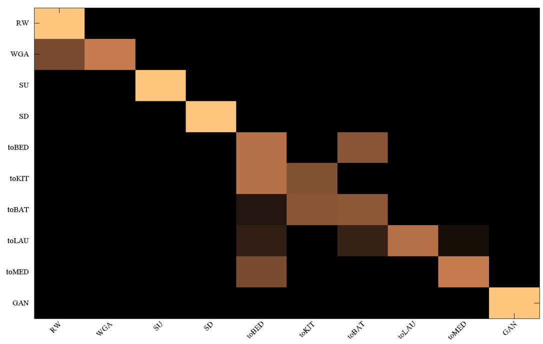


(a) Confusion Matrix of ADLs

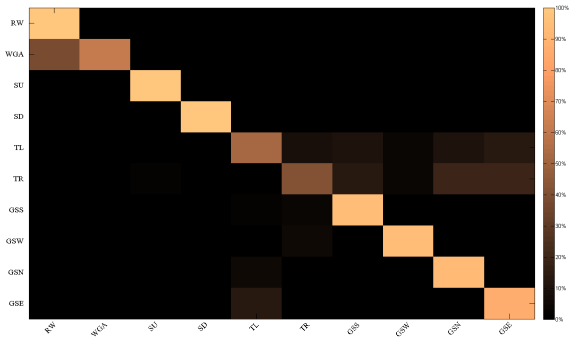


(b) Confusion Matrix of APs

Figure 4.7: Confusion Matrix of ADLs and APs inferred by the HHMM Model for Walker user



(a) Confusion Matrix of ADLs



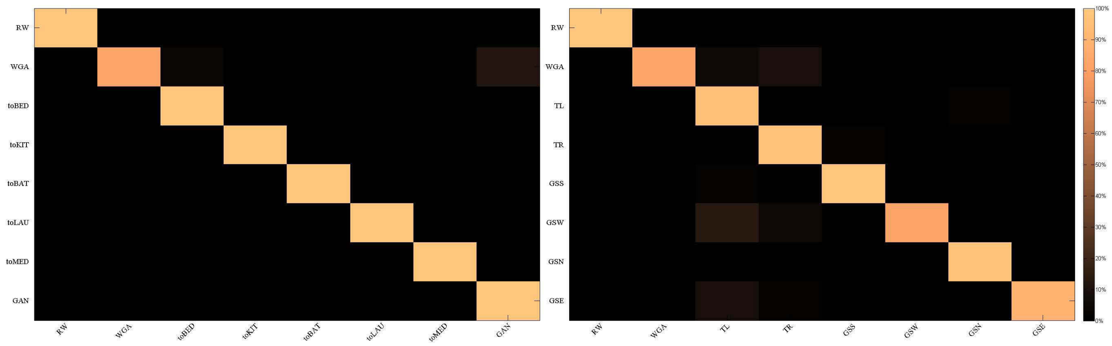
(b) Confusion Matrix of APs

Figure 4.8: Confusion Matrix of ADLs and APs inferred by the L-DBN Model for Walker user

65.99% and 64.30% for the L-DBN framework.

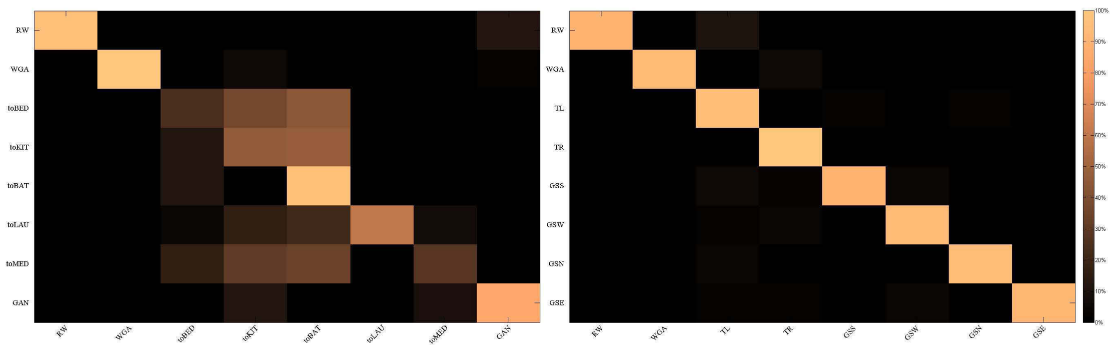
In the HHMM framework, the APs were inferred at the intermediate level (Level 2) with an accuracy of 88.88% and 93.73% for the walker and wheelchair platforms respectively. The same were inferred with an accuracy of 86.61% and 92.83% for the walker and wheelchair platform respectively at Level 1 using the L-DBN framework. The confusion matrix of ADLs and APs for both the framework and both the platforms is depicted in Figures 4.7, 4.8, 4.9 and 4.10.

The APs were inferred with very similar accuracy for both the HHMM and L-DBN frameworks, however the ADLs inferred by the L-DBN framework were significantly reduced as compared to the HHMM framework. This further validates the suggestion



(a) Confusion Matrix of ADLs (b) Confusion Matrix of APs

Figure 4.9: Confusion Matrix of ADLs and APs inferred by HHMM Model for wheelchair user



(a) Confusion Matrix of ADLs (b) Confusion Matrix of APs

Figure 4.10: Confusion Matrix of ADLs and APs inferred by L-DBN Model for wheelchair user

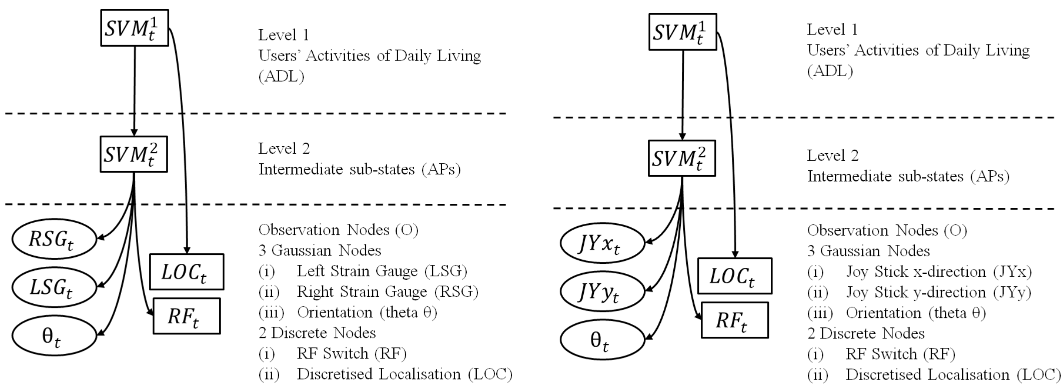
that the proposed hierarchical structure is able to take full advantage of the spatio-temporal nature of ADLs.

4.6.1 Comparison with Discriminative Model

We also compared the results of HHMM framework with a Support Vector Machine (SVM) based discriminative classifier. SVM have become one of the most popular classification methods in the Machine Learning field in recent years finding its applicability in various real world problems such as activity recognition, text classification, bio-informatics and many more. An SVM uses a discriminant hyper-plane to demarcate between different classes such that the hyper-plane maximises the margins, i.e. the distance from the nearest training point [Vapnik, 1998]. In this chapter we used a two level SVM classifier (Figure 4.11). The SVM classifier at Level 2 (SVM^2) classifies the APs using the raw sensor data which is then fused with discretised localisation information used by the SVM on Level 1 (SVM^1) to predict the overall ADL. Since SVM learning is supervised, classification accuracy of APs was high, in the order of 98% for the walker and 97% for the wheelchair, as the atomic classes are more easily differentiable. However the temporal constraints in the overall ADL classification cannot be so easily encapsulated by maximum margin approaches such as SVM, and accuracies of around 79.93% and 75.06% were attained for the walker and wheelchair platform respectively. The ADLs and APs inferred by HHMM, L-DBN and 2-stage SVM classifier is listed in Table 4.4 and 4.5 and shown in Figure 4.12 and 4.13 for the walker and wheelchair platform respectively.

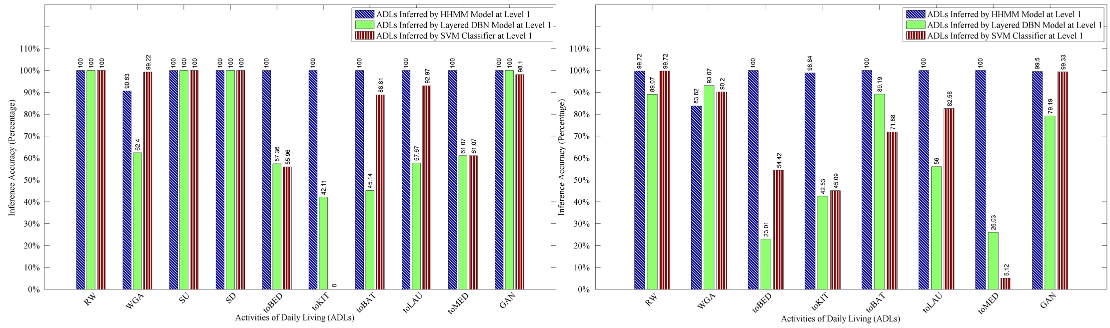
4.6.2 Inferring ADLs and APs using On-line Inference

As detailed in Chapter 3, the primary aim of inferring user ADLs and the associated APs is to develop a model that can provide assistance to the user as and when required. To this end, the ADLs and the associated APs need to be inferred in real time, hence on-line inference was performed using the forward algorithm described in Chapter 2. The on-line inference was done only for the HHMM framework, given the lower inference



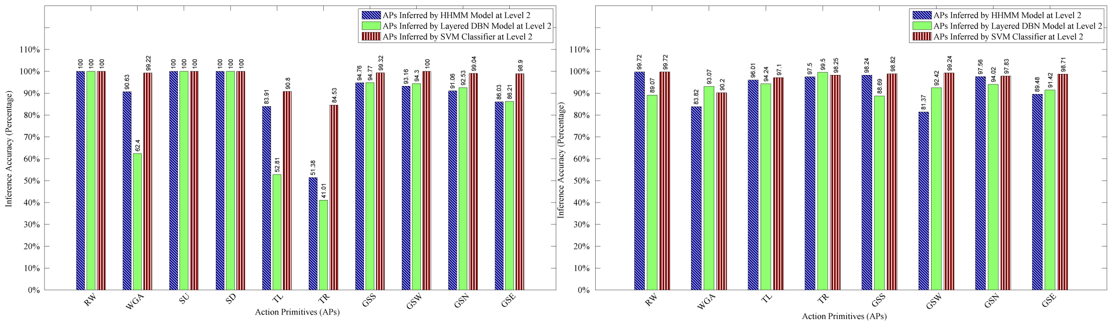
(a) 2 level SVM Model for Walker user (b) 2 level Model used for Wheelchair user

Figure 4.11: 2 Level SVM Model used to infer ADLs of walker user (Figure (a)) and wheelchair user (Figure (b))



(a) Comparison of ADLs for walker user (b) Comparison of ADLs for wheelchair user

Figure 4.12: Comparison of ADLs inferred by HHMM, L-DBN and SVM Classifier for walker and wheelchair user

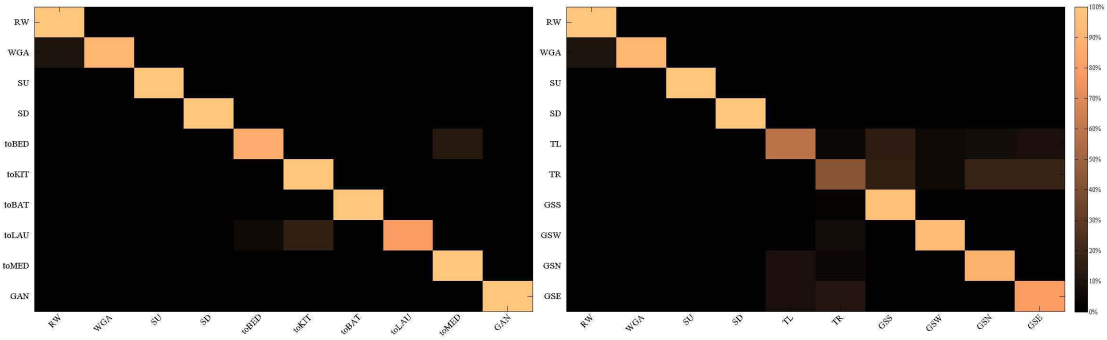


(a) Comparison of APs for walker user (b) Comparison of APs for wheelchair user

Figure 4.13: Comparison of APs inferred by HHMM, L-DBN and SVM Classifier for walker and wheelchair user

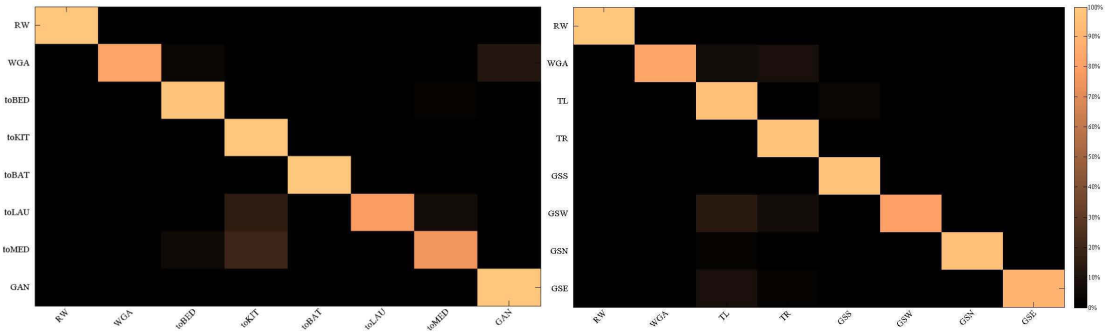
accuracy achieved by the L-DBN model already using off-line inference.

ADLs were inferred with an overall accuracy of 91% for both the platforms, whereas inference accuracy for APs was 86% and 93% for the walker and wheelchair platforms respectively. The confusion matrix of ADLs and APs inferred using the on-line forward algorithm for the walker and wheelchair platform are shown in Figures 4.14 and 4.15. Furthermore, a comparison of the inference accuracies achieved using the off-line and on-line inference engines are shown in Figures 4.16(a) and 4.16(b). On-line inference accuracy is reduced by around 10% when compared to the off-line method, which is not surprising given the reliance of the current user state only on observations till the current time, and the past state. The probabilistic inference confidence of the user performing the activity of going to the laundry room from the bedroom is shown in Figure 4.17.



(a) Confusion Matrix of ADLs (b) Confusion Matrix of APs

Figure 4.14: Confusion Matrix of ADLs and APs inferred by HHMM model using on-line inference for walker user

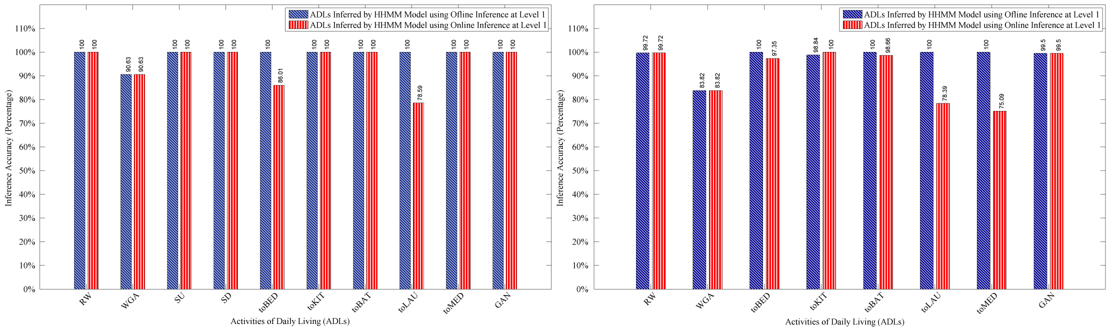


(a) Confusion Matrix of ADLs (b) Confusion Matrix of APs

Figure 4.15: Confusion Matrix of ADLs and APs inferred by HHMM model using on-line inference for wheelchair user

4.7 Summary

In this chapter we presented a mechanism to model ADLs based on intrinsic human motion as captured by the physical sensor present on the walker and wheelchair platforms. The pool of APs used to model ADLs consisted of human motion patterns while the individual performed different ADLs using ambulatory robots. This was a further improvement to the approach used in the previous chapter, as the information related to junction point based topological map was not required to model ADLs. The HHMM framework proves to be a powerful tool to model ADLs, as it inferred APs with accuracies in the range of 88% and 92% respectively for the walker and wheelchair platform, and rather significantly ADLs were predicted with an accuracy of 99% and 98% for the



(a) Comparison of ADLs for walker user (b) Comparison of ADLs for wheelchair user

Figure 4.16: Comparison of ADLs and APs inferred by HHMM framework using off-line and on-line inference for walker and wheelchair user

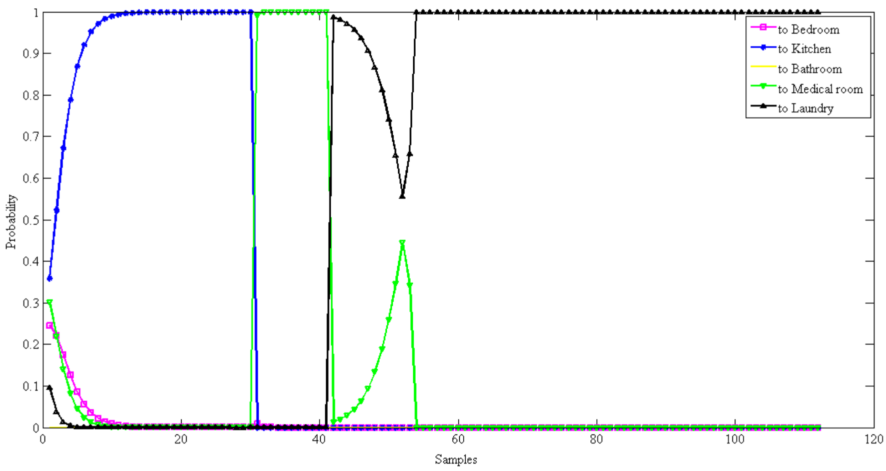


Figure 4.17: Probability evolution of inferring ADL of going to the Laundry from Bedroom. Note that the inference is performed using the observation available till the current time.

same platforms.

The inference accuracies of the HHMM model were also compared with those of a L-DBN model and a 2-stage SVM classifier. The L-DBN model and staged SVM model were used so as to resemble the modelling characteristics of the HHMM model, where APs are inferred at the lower level and are then combined in different sequence to infer the ADLs. The APs are inferred with around the same accuracies, yet inference of the ADLs reduced substantially for the L-DBN and SVM models, which again proves that modelling ADLs at multiple levels plays a critical role in predicting ADLs, while an L-DBN or a 2-stage SVM are not capable of exploiting these relationship. Lastly, we also compared the results of two different ways of inferring user ADLs, i.e. inferring the ADLs performed by the user from the entire observation sequence (off-line inference) and inferring ADLs at a given point in time based on available observations till that time (on-line inference).

So far, we have employed various tactics to model ADLs performed by the user with the support of ambulatory robots such as a power walker and a robotic wheelchair. We exploited different sensor information that can be used to extract meaningful information to model these ADLs. In the next chapter we extend the usage of our proposition of modelling ADLs which are related to grasping and manipulation of everyday objects.

Chapter 5

Modelling Grasping and Manipulation Activities

5.1 Introduction

In Chapter 3 and 4, we presented different techniques which can be used to model Activities of Daily Living (ADLs) performed by users of ambulatory robots such as a power walker and a robotic wheelchair. The Hierarchical Hidden Markov Model (HHMM) framework demonstrated to be a powerful tool to model complex ADLs from low level sensor information. As per the report published by World Health Organisation [World Health Organisation, 2004], apart from locomotion related ADLs, activities such as eating, bathing and toileting are also defined under the umbrella of basic ADLs. These ADLs involve grasping and manipulation of different objects such as a mug, bottle, brush, soap, etc. In this chapter we focus on modelling ADLs performed by users which are related to grasping and manipulation of everyday objects.

We extend the usage of the HHMM based temporal model, and demonstrate how it can be used for representing and learning object grasping and manipulation activities. The model builds upon a dictionary of APs which are combined to compose and describe complex ADLs. The hierarchical nature of the framework allows typical activities to

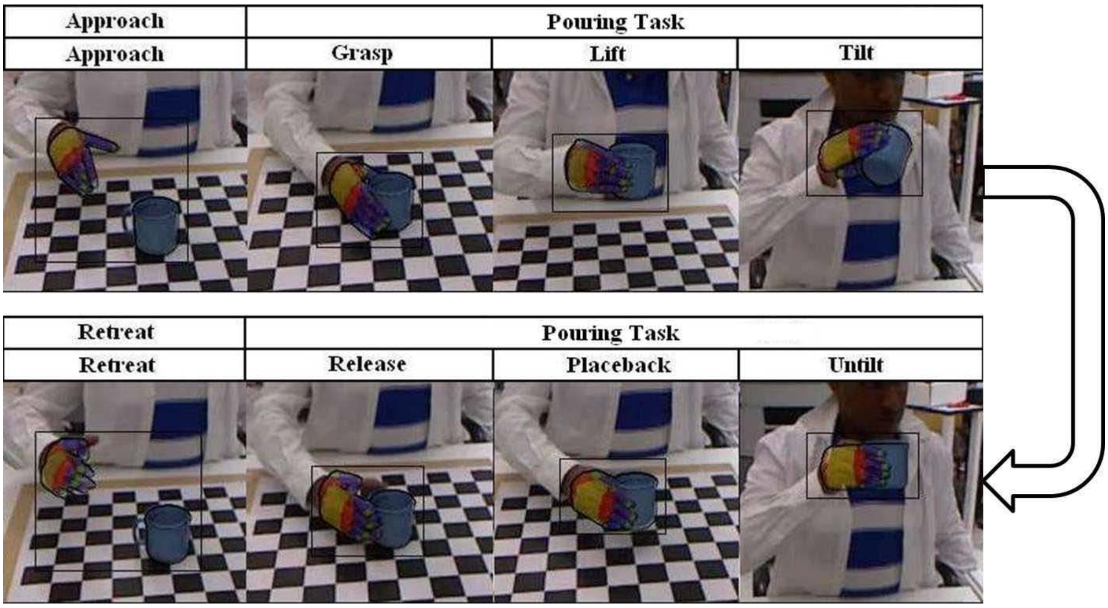


Figure 5.1: Activity of *Pouring* water from mug subdivided into action primitives. Each image depicts the output of hand-object tracking algorithm.

be decomposed into different APs which are learned by the model at different levels of hierarchy. The example shown in Figure 5.1 decomposes a pouring activity into sequence of APs. The APs provide the necessary tool to describe a complex activity as a sequential combination similar to a natural language description. The proposed framework is capable of learning this at different levels i.e. the APs are learned and inferred by observing the hand-object interactions and their motion in the Cartesian space, whereas ADLs are inferred by learning the sequence of APs. We also compare the inference accuracy of the HHMM model with that of a HHMM/SVM hybrid model.

5.2 Related Work

A challenge in modelling grasping and manipulation activities is the extraction and representation of these activities from the raw sensory data. Given the inherent level of uncertainty and noise in this data, it is difficult to model these activities in a deterministic manner. Different probabilistic techniques have been applied by researchers to model grasping and manipulation activities [Khansari-Zadeh and Billard, 2010], [Dindo and Schillaci, 2010], [Pastor et al., 2009], [Nemec and Ude, 2012], [Krüger et al., 2010].

Learning by imitation is an approach that has been used by roboticists for bootstrapping learning of robot activities based on human observation. Preliminary work done by Ijspreet and colleagues used a Control Policy (CPs) based approach to represent complex dynamical systems based on human movements [Ijspeert et al., 2002]. These CPs, which represent various human like movement plans, are derived based on ease of representation, compactness, robustness against changes in the dynamic environment, re-usability and overall simplicity in learning different human movement trajectories. This Dynamic Motion Primitive (DMP) based framework was later illustrated in a number of applications related to humanoid robotics which involved planning, movement recognition, perception-action coupling, imitation and general reinforcement learning [Schaal et al., 2004]. Khansari-Zadeh and Billard [Khansari-Zadeh and Billard, 2010] used a learning method called Stable Estimator of Dynamical Systems (SEDS) to learn the parameters of a time invariant dynamical system to ensure that all motions closely follow the demonstrations while ultimately reaching and stopping at the target. Dindo and Schillaci [Dindo and Schillaci, 2010] used an imitation learning based approach to recognise the skills being observed and reproduce them using a generative model. They utilised a Growing Hierarchical Dynamic Bayesian Network (GHDBN) based generative model capable of learning various skills at different levels of hierarchy and also able to adapt (learn new skills) as new observation sequences are available. The model learned and reproduced three actions i.e. *Dislocate, Approach* and *Hit*. Pastor *et. al.* [Pastor et al., 2009] used a Dynamic Movement Primitive (DMP) framework in which the recorded movements were represented using non-linear differential equations. The movement library consisted of actions such as *grasping, placing* and *releasing*. Nemec and Ude [Nemec and Ude, 2012] in their recent work also used a DMP based system to represent primitive movements. The DMP library used in their experiment consisted of activities like *reaching, pouring, wiping, shaking, cutting, power grasps etc.* A Parametric Hidden Markov Model (PHMM) was proposed by Krüger *et.. al.* to represent various action primitives [Krüger et al., 2010]. The framework was trained in an unsupervised manner and synthesized movement trajectories as a function of their desired effect on the object. The set of actions learned were *approach, grasp, push forward, push side, move side, rotate* and *remove*. Song *et. al.* used structure learning to exploit

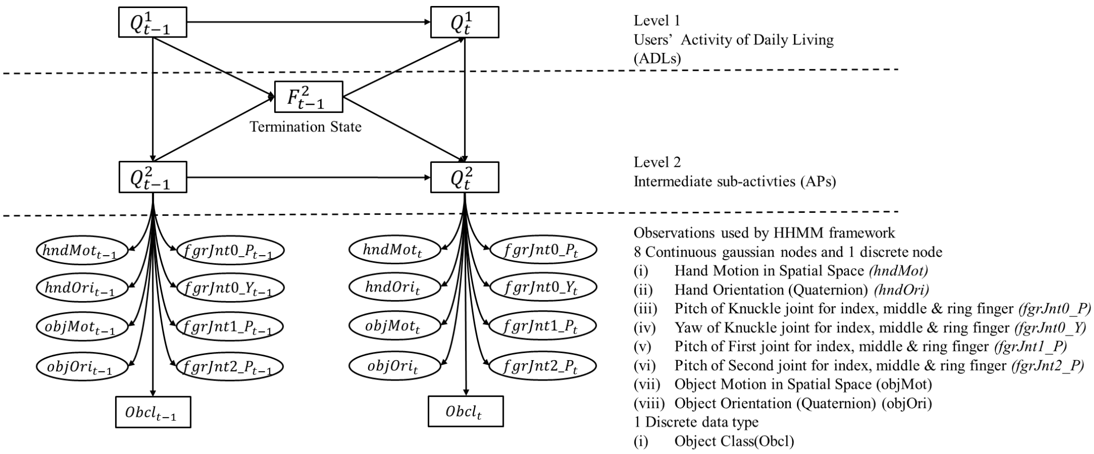


Figure 5.2: HHMM Model used to infer action primitives and long term user activities using hand and object features (described in Table 5.1)

the dependencies between hand and object to generate the structure of a Bayesian Networks (BN) [Song et al., 2011b], [Song et al., 2011a]. The evolved structure was used to predict the activity performed by the user based on the type of action, and the object being manipulated. However, these activities were predicted based on grasp instances, and the prediction process did not exploit features from the entire trajectory that was followed by the arm when performing the given activity.

The approach used in this thesis is novel for two reasons: firstly, the entire activity sequence is clustered into a pool of different APs and secondly, the unified probabilistic framework exploits spatial relationships to learn both APs and the time dependent relationship between them, so as to be able to accurately predict complex manipulation activities at the highest level of abstraction. In real-time applications, clustering activities into different APs becomes an important criterion as the time taken by any user to perform a given activity will vary (even for the same user), suggesting a high variability in users remaining within a given (action primitive) state. In order to accommodate this variability, the use of hierarchical models with specific conditions to model the end of sub-processes is an important proposition. However, considering a unique user state at each time instance is computationally intractable as the state space would grow unbounded.

5.3 Modelling ADLs using Probabilistic Models

A HHMM based probabilistic framework similar to that described in Chapter 3 and 4 was used to model grasping related ADLs. As illustrated in Figure 5.2, the top level of the framework inferred ADLs, the intermediate level inferred APs while the lowest level corresponds to the features of object-hand interaction in the Cartesian space.

5.3.1 Hierarchical Hidden Markov Model

The 2-level HHMM framework used to model grasping and manipulation related activities can be represented using a Hierarchical Dynamic Bayesian Network (H-DBN) framework as shown in Figure 5.2. The ADLs were decomposed into APs (listed in Table 5.2), which were based on visual inspection of the data. The H-DBN framework consists of three types of nodes: Q_t^d, F_t^d, O_t. The ADLs are represented by Q_t^1 whereas the APs are represented by Q_t^2. Given the parameters (Q_t^d, O_t, F_t^d), the H-DBN defines the joint distribution over the set of variables that represents the evolution of the stochastic process over time. These distributions are in the form of prior distributions (initial probabilities of the state variables at each level), the transition probabilities and the observation probabilities. The prior probabilities, transition probabilities and termination probabilities are defined similar to that in Chapter 3 and 4 and the definitions are repeated here for convenience. The prior probabilities at both the Level 1 and 2 is given by Equation 5.1 where as the transition probabilities at level 1 and 2 are given by Equation 5.2 and 5.3 respectively. The termination probabilities are given by Equation 5.4. The observation nodes are modelled as both Gaussian and discrete. The CPDs for Gaussian and discrete nodes is given by Equation 5.5.

$$\begin{aligned} P(Q_1^1) &= \pi^1(j) \\ P(Q_1^2) &= \pi_k^2(j) \end{aligned} \tag{5.1}$$

$$P(Q_t^1 = j | Q_{t-1}^1 = i, F_{t-1}^2 = f) = \begin{cases} A^1(i,j) & \text{if } F_{t-1}^2 = 0 \\ \pi^1(j) & \text{if } F_{t-1}^2 = 1 \end{cases} \tag{5.2}$$

$$P(Q_t^2 = j | Q_{t-1}^2 = i, F_{t-1}^2 = f, Q_t^1 = k) = \begin{cases} A_k^2(i,j) & \text{if } F_{t-1}^2 = 0 \\ \pi_k^2(j) & \text{if } F_{t-1}^2 = 1 \end{cases} \tag{5.3}$$

$$P(F_t^2 = 1 | Q_t^1 = k, Q_t^2 = i) = A_k^2(i, end) \tag{5.4}$$

In Equation 5.1, 5.2, A^1 and π^1 represent the transition and initial probabilities respectively at Level 1 whereas in Equation 5.2 and 5.4, A_k^2 and π_k^2 represents the same at Level 2 given the state at Level 1 is k.

$$\begin{aligned} P(O_t | Q_t^1 = i) &= N(\mu_i, \Sigma_i) \\ P(O_t | Q_t^1 = i) &= C(i) \end{aligned} \tag{5.5}$$

In everyday life, a single object can be used to perform many activities; for example, a mug can be used for drinking, pouring, or can be handed over to another person. For each ADLs, the user approaches to grasp the object, carries out the desired activity, and then retreats the hand after releasing the object. The actions of approaching and retreating the hand occur whenever the object is used, and cannot be described as part of the activity sequence, as it is only the specific activity itself that uniquely characterises an ADL. Hence such APs, e.g. approaching to grasp an object (**APPRH**), and retreating after the object is released (**RETRT**), are not defined as components of any ADLs listed in Table 5.3, but are treated as independent ADLs in themselves. Such ADLs, are inferred at both levels in the HHMM framework. To better illustrate this concept, consider the example in Figure 5.1.

The specific activity of pouring cannot be inferred until the time the mug has been grasped. Hence when the user first approaches to grasp the mug, the ADL and the AP remain the same and are defined at both levels. Once the object is grasped, the ADL can be inferred based on the type of grasp and the object. Hence, the HHMM model will infer the ADL of *Approach* at both the levels whereas *Pour* is inferred at the higher level (1) and the corresponding subsequence of APs *Grasp from Middle* $\Rightarrow$ *Tilt* $\Rightarrow$ *Untilt* $\Rightarrow$ *Put Back* $\Rightarrow$ *Release* is inferred at the lower level (2). Similarly after releasing the

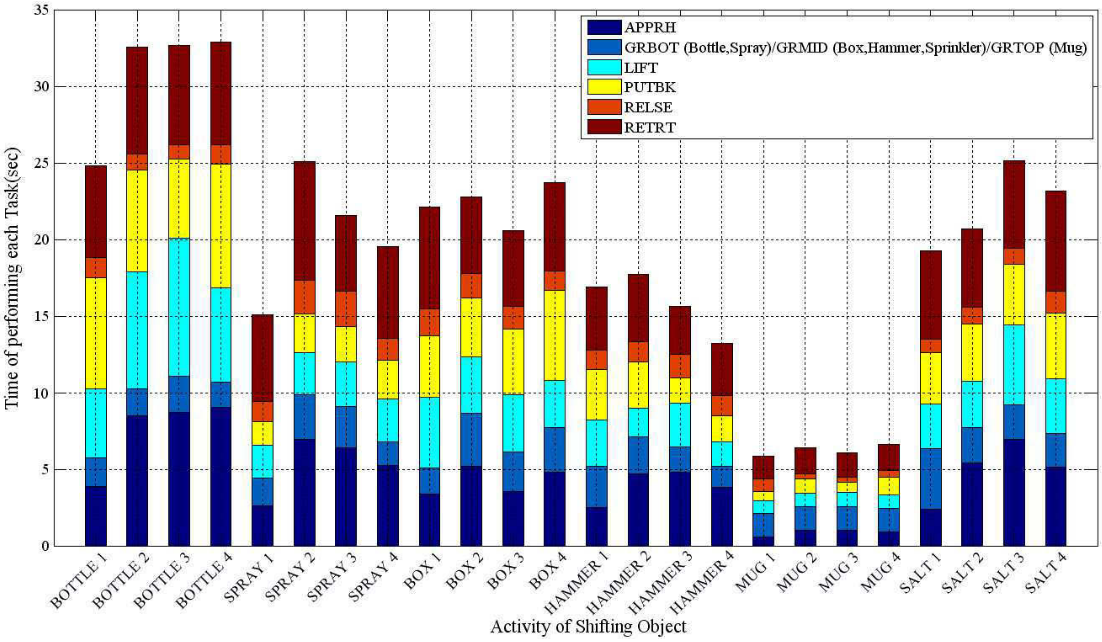


Figure 5.3: Time taken by each action primitive (APs) to perform the activity of shifting objects. Note that the time taken for shifting the same object and the time spent within each AP varies between same and different objects

object the ADL *Retreat* is inferred at both levels.

The hand and object features used at the observation level, are extracted using a hand-object tracking algorithm (details are given in Section 5.4), which represents the interaction between the hand and object and its movement in Cartesian space.

Feature	Dim.	Description
$hndMot$	3	Hand motion in cartesian space
$hndOri$	4	Hand orientation (quaternion)
$fgrJnt0\_P$	1	Pitch of knuckle joint for index, ring & middle finger
$fgrJnt0\_Y$	1	Yaw of knuckle joint for index, ring & middle finger
$fgrJnt1\_P$	1	Pitch of first finger joint for index, ring & middle finger
$fgrJnt2\_P$	1	Pitch of second finger joint for index, ring & middle finger
$objMot$	3	Object motion in cartesian space
$objOri$	4	Object orientation (quaternion)
$Obcl$	6	Object class

Table 5.1: Hand and object features used by the HHMM framework

Action Primitive (APs)	Abbrev.	Description
Approach	APPRH	Approach to grasp objects in a given space
Approach with twisted hand	APTWH	Approach to grasp objects with inverted hand
Retreat	RETRT	Retreat hand into original position
Putback	PUTBK	Place back the grasped object
Grasp from top	GRTOP	Grasp object from top
Grasp from handle	GRHDL	Grasp object from handle (if any)
Grasp from middle	GRMID	Grasp object from middle
Grasp from tool use end	GRTUE	Grasp object from tool use end
Lift object	LIFT	Lift grasped object
Tilt object	TILT	Tilt grasped object
Un-tilt object	UNTLT	Un-tilt grasped object
Lower object (tool)	LWRTL	Lower object for usage
Raise object (tool)	RAITL	Raise object for usage
Move object towards You	MVTOU	Move object towards you
Release	RELSE	Release the grasped object
Grasp from bottom	GRBOT	Grasp object from bottom
Invert object	INVRT	Invert the grasped object by 180 degrees
Press and release trigger	PERLTGR	Press and release trigger of spray bottle
Shake salt sprinkler	SHAKE	Shake salt sprinkler to sprinkle salt

Table 5.2: Action Primitives to perform various activities

5.4 Data Collection

In order to validate our proposed approach, we collected data using an RGB-D kinect sensor while human subjects demonstrated grasping and manipulation activities. The parameters that describe the configuration of the users' hand and the configuration of the object while performing the activities had to be extracted from the 3D video stream data. The extracted features which involved the interaction between the hand and object had to be such that they could be mapped to the motion of a robotic arm for activity synthesis/imitation. In order to extract such information we combined the methods presented in [Oikonomidis et al., 2011a] and [Oikonomidis et al., 2011b] towards a system that can track both the hand and object while they are interacting in Cartesian space. The hand tracked using the technique described in [Oikonomidis et al., 2011b], which optimizes the objective function that quantifies the discrepancy between a hypothesis over the scene state and the actual observations. The tracking algorithm also accommodated tracking of the object and its motion in Cartesian space. At each new frame a new tracking optimization was performed that was initialized in

ADLs	Abbrev.	Description
Approach	APPRH	Approach to grasp objects in a given space
Approach with twisted hand	APTWH	Approach to grasp objects with inverted hand
Retreat	RETRT	Retreat hand into original position
Pour	POUR	Activity of pouring from a mug or bottle
Handover	HNDOVR	Activity of handing over an object to another person
Tool Use (Hammer)	TLUSE	Hammering a nail
Spray	SPRAY	Spraying from a spray bottle
Dish Wash	DSHWSH	Loading an object like a mug in a dishwasher
Drink	DRINK	Drink from a mug or bottle
Shift	SHIFT	Shift object for a one location to another
Sprinkle Salt	SPRINKLE	Sprinkle salt using a salt sprinkler

Table 5.3: Users' Activities of Daily Living (ADLs)

the vicinity of the solution for the previous frame. The reference 3D coordinate system was conveniently defined to reside on the demonstration table (seen in Figure 5.1), which became a chess-board calibration pattern. All objects used in the manipulative activities were painted blue, as per Figure 5.4, so as to rely upon a single, uniform appearance model for tracking, thus facilitating the overall set-up.

To initialise the hand and object position we employed a similar technique to the one specified in [Oikonomidis et al., 2011a], [Oikonomidis et al., 2011b] and [Papazov and Burschka, 2011]. To successfully track the hand, the tracking algorithm expected the hand to be at a given initial position in the space. To initialise the pose of the object, we integrated the tracking algorithm with the RGB-D based registration method used by Papazov [Papazov and Burschka, 2011].

The features extracted from the experimental results to validate the proposed work are listed in Table 5.1. They consist of the 3D motion (translation and rotation) of the hand and the object being manipulated. The features in the data also include a selection of the rotational joint movements of three of the fingers, index, middle and ring. The derived trajectory provided information about the motion of the hand and object, whereas the rotational motion (yaw, pan, tilt) added information about their corresponding orientation in space. Furthermore, the movement of the finger joints provided details about the grasping of the objects. All these data features were utilised to predict the APs at the lower level.

It is worth noting that the primary goal in this work is the representation of human

Figure 5.4: Objects used to perform manipulation activities

grasping and manipulation so that these behaviours can effectively be learned from a human teacher and ultimately transferred to a robot arm. Kinematic models and degree of freedom (DOFs) between a human arm and a robotic manipulator differ, thus the paths followed by both in exercising a manipulation activity will diverge. However, for a capable anthropomorphic arm the interactions between a robotic arm and the objects in their surroundings (e.g. grasping the object with a particular pose in order to accomplish the desired activity) will be of similar nature - subject of course to their differing kinematic arrangements. As such, the APs learned by the robot (*GRTOP, TILT* etc.) and the sequences needed to accomplish a given task are directly transferable to any grasping manipulator of sufficient dexterity.

5.5 Results

To test the proposed methodology, we used a selection of objects used in everyday life. We intentionally selected objects that can be used in the context of more than one activity, e.g. a mug and a bottle which can be used both for drinking and pouring. We selected the six objects depicted in Figure 5.4 to perform the ADLs listed in Table 5.3. Data were collected for a single user, who repeated the same activity 4 times to capture variations which might occur in performing the same activity. The user was asked to

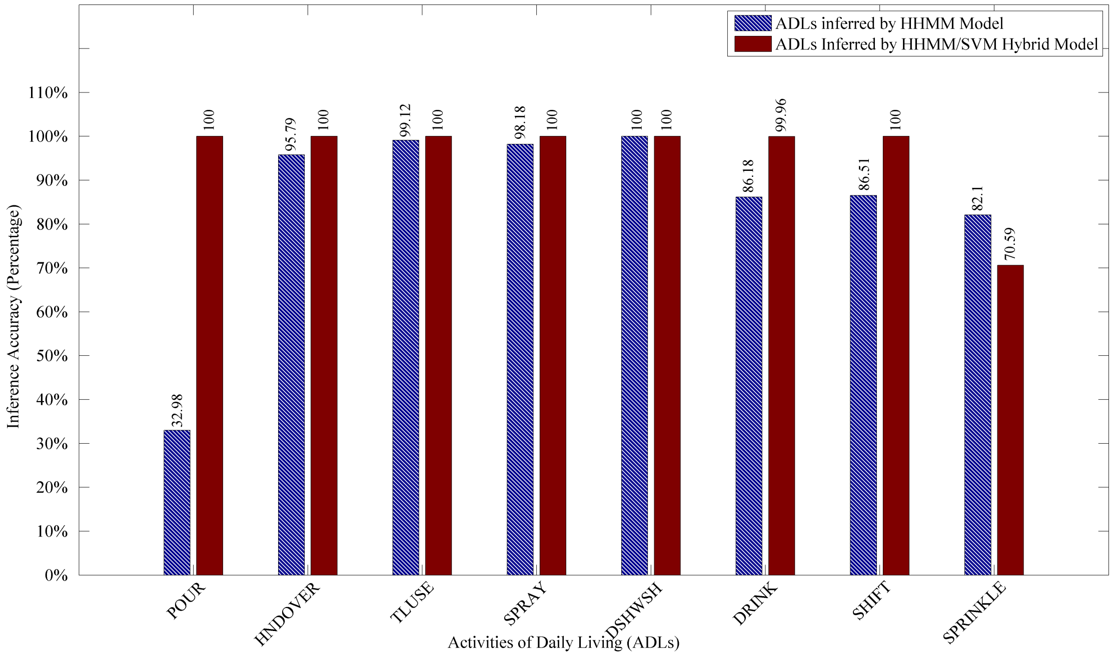


Figure 5.5: Comparison of ADLs inferred by HHMM and HHMM/SVM Hybrid Models

perform each activity such that it resembles natural execution. The video and depth data were collected at a rate of 30 frames per second. The motion of hand and object was extracted off-line using the hand-object tracking algorithm described in Section 5.4. The output of the tracking algorithm provided data for orientation and motion of hand and object motion in the cartesian space. The tracker also extracted the features for each finger joint. All the activities were decomposed into a total of 19 interpretable APs based on visual inspection, and are collected in Table 5.2. It is important to emphasize that each AP represents a cluster, which is a continuous, time-varying trajectories of the feature set, and not a single instance.

Due to the time variation in performing different activities, the time spent in executing each AP will vary. This would be the case even if it is the same activity that is repeated over and over again. To illustrate this, Figure 5.3 shows an example of the time taken to perform the activity of SHIFT which involved shifting different objects from one location to another. It can be seen that the time taken for each AP in a given activity varied even if the activity was repeated on the same object. For example, when comparing the activity of shifting a bottle, *BOTTLE 1* took significantly less time than the other three

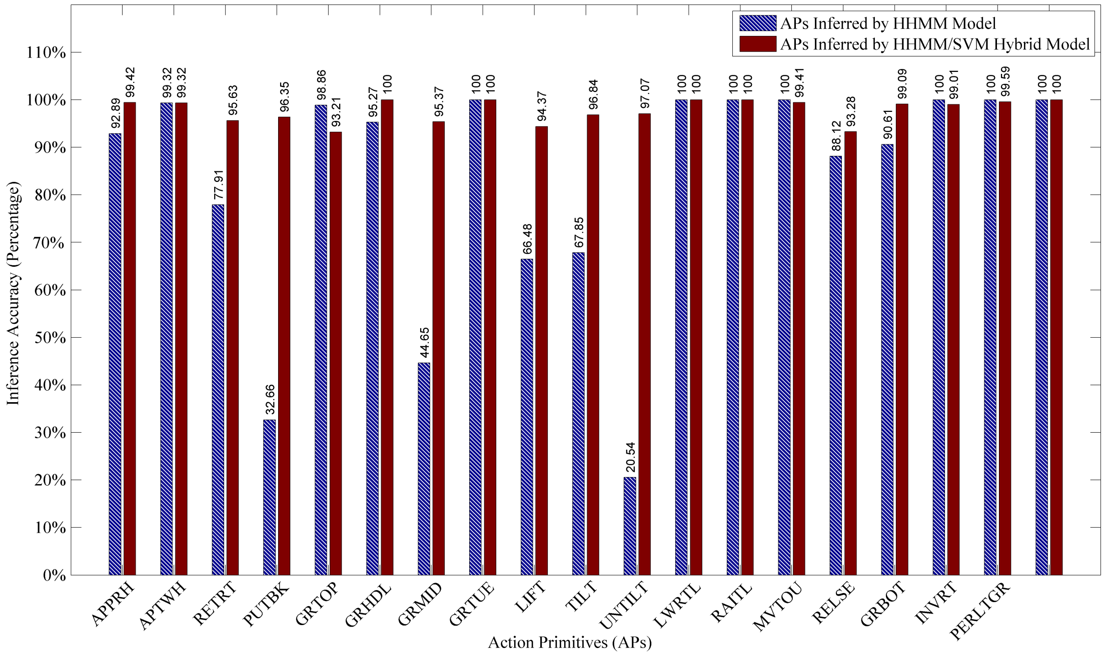


Figure 5.6: Comparison of APs inferred by HHMM and HHMM/SVM Hybrid Models

times (*BOTTLE 2, BOTTLE 3, BOTTLE 4*). This variation in the time required to complete the same activity on different occasions was reflected in/by the time taken to undertake each AP.

The HHMM model (Figure 5.2) was trained and tested using the captured hand and object motion data, as described in Section 5.4. The data set was manually labelled for both APs and ADLs for cross validating the inference accuracy. We divided the data set into two equal halves for training and testing purposes. We used the BNT toolbox [Murphy, 2002] to learn and infer APs and ADLs using the proposed HHMM model. Expectation Maximisation (EM) was used to learn APs and ADLs, and the Maximum Likelihood Estimator was used for inference. The features used by the HHMM framework, including their dimensions are listed in Table 5.1.

The APs were inferred with an overall accuracy of 72% at the intermediate level (Level 2) of the HHMM model whereas the ADLs were inferred with 86% accuracy (at the higher level). The inference accuracy of predicting each AP and the ADLs are graphically depicted in Fig 5.6 and 5.5 respectively.

Most of the APs were inferred with greater than 72% accuracy. APs such as putback

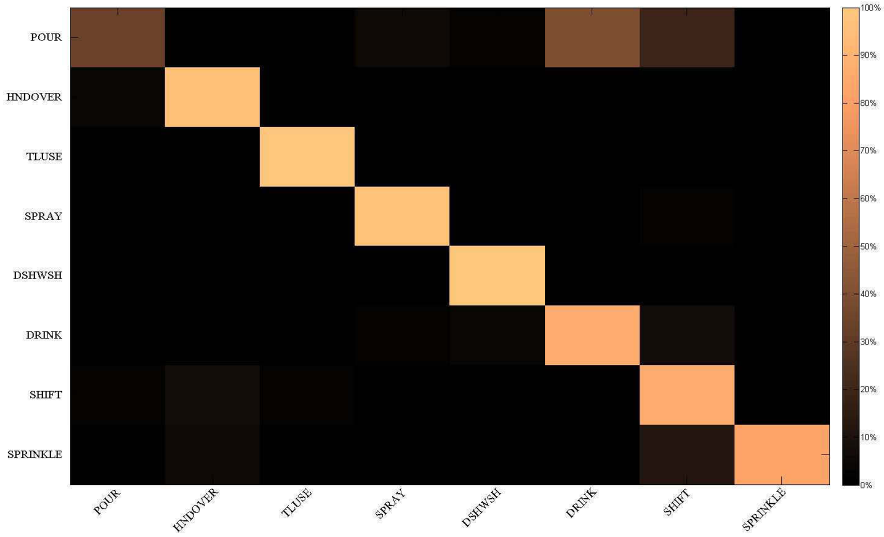


Figure 5.7: Confusion Matrix for ADLs inferred by HHMM Model at Level 1

(**PUTBK**), tilt (**TILT**), un-tilt (**UNTLT**), grasp object from middle (**GRMID**) and lift (**LIFT**) were inferred with less than 70% accuracy. **PUTBK** is often confused with **LIFT** (as can be seen in Figure 5.8) as both APs follow almost the same trajectory in Cartesian space. A very high level of confusion is also evident between action states **TILT** and **UNTLT**. This is not surprising as in the continuous space both these actions are performed one after another, and hence the framework is unable to clearly discriminate between them. Lastly, a high level of confusion exists between the state of grasping the object from middle and bottom due to the unavailability of relevant information such as distance offset between the center of object and grasping points.

At a higher level, with the exception of the activities **POUR** and **DRINK**, all other activities were inferred with fairly high accuracy (refer to the confusion matrix in Figure 5.7). Confusion occurred between these two activities as there is minimal difference in the sequence of APs followed to perform both drinking and pouring.

5.5.1 Comparison with HHMM/SVM Hybrid Model

We also compared the accuracy of the HHMM model with that of a hybrid HHMM/SVM model. As described in Chapter 3, HMM/SVM hybrid model has been successfully used

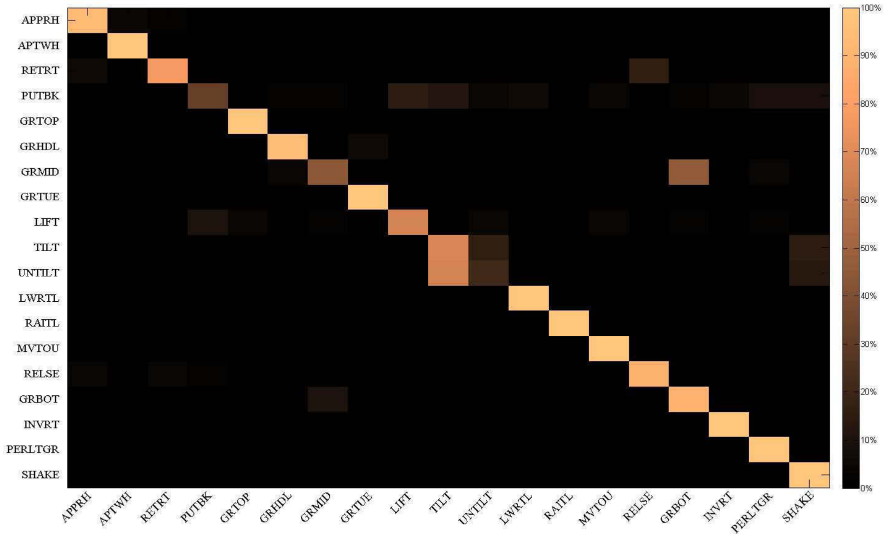


Figure 5.8: Confusion Matrix for APs inferred by HHMM Model at Level 2

in a number of application such as automatic speech recognition [Stadermann and Rigoll, 2004], tele-operation [Castellani et al., 2004] and modelling of facial action temporal dynamics [Valstar and Pantic, 2007]. Stadermann used a SVM/HMM hybrid model for speech recognition which combines the strong classification capabilities of the SVM with the time varying modelling capability of the HMM framework [Stadermann and Rigoll, 2004]. Valster and Pantic also exploited the capabilities of SVM/HMM hybrid model for facial action recognition. In this application the SVM classifies the distinction between temporal (facial expression) phases at a single point in time which are then combined over a time period by the HMM model to predict the temporal dynamics [Valstar and Pantic, 2007]. A similar technique was used by Castellani and colleagues for analysing and segmenting various teleoperation activities [Castellani et al., 2004]. In all these approaches, the capability of the SVM to handle non-linear data through kernel induced feature maps is exploited, which in turn can be used by the HMM to model the temporal relationship between data points.

We used an SVM classifier to predict the APs at a single time instance which were then combined in a temporal space within the HHMM model to predict high level activities. The HHMM/SVM hybrid model used for comparison is shown in Figure 5.9. To make the comparison fair, we used a HHMM framework instead of a flat HMM model so that

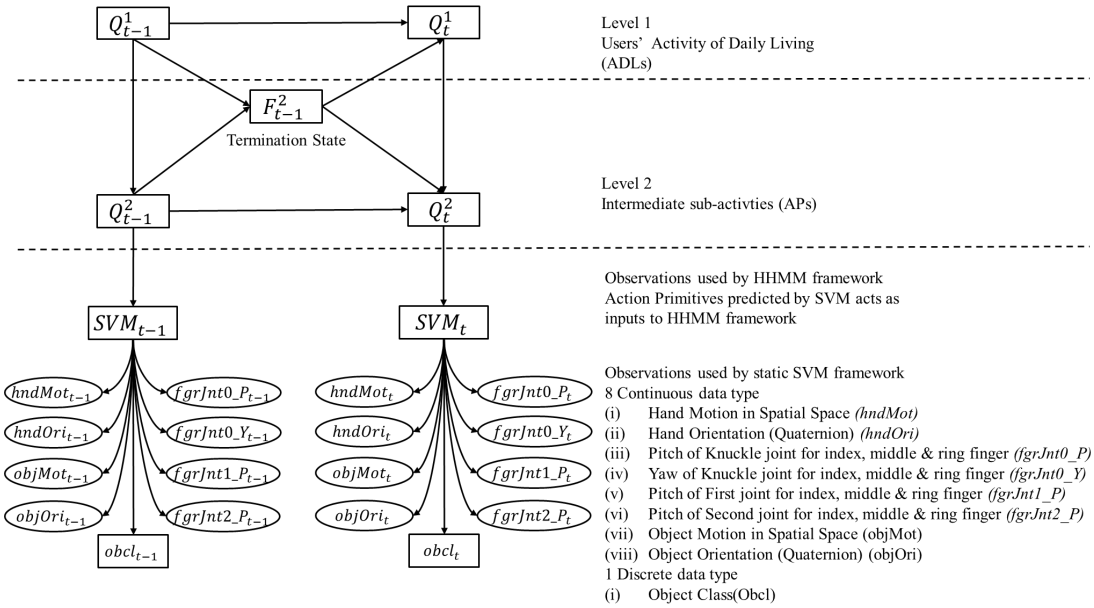


Figure 5.9: HHMM/SVM Hybrid Model used to infer action primitives and long term user activity using different hand and object features. The SVM classifier at the lower level classifies action primitives using hand and object features which are then used by the HHMM framework to predict the long term activities.

the self transition and inter state transition characteristics at Level 1 and 2 remained the same for both models. The high level activities were inferred at Level 1 with an overall inference accuracy of 95% (Figure 5.5). The APs were inferred with an overall accuracy of 97% at Level 2 (Figure 5.6), which corresponds to a direct mapping of the APs classified by the SVM model. Further, the features used for the SVM classifier at each time step were the same as those used by the HHMM model.

Most of the APs were inferred with approximately the same accuracies for both HHMM and HHMM/SVM hybrid models except for **PUTBK, GRMID, LIFT, TILT, ULTILT** (confusion matrix for inferring both ADLs and APs is shown in Figure 5.10 and Figure 5.11) . The HHMM model is less able to discriminate between these classes as described in Section 5.5. However, the SVM is able to predict these APs with high accuracy, which is not surprising as SVM possess strong capability to discriminate between these classes with minimal difference in observation. The HHMM/SVM hybrid model outperformed the HHMM model in inferring ADLs given the strong classification of APs by the SVM classifier as compared to the HHMM model.

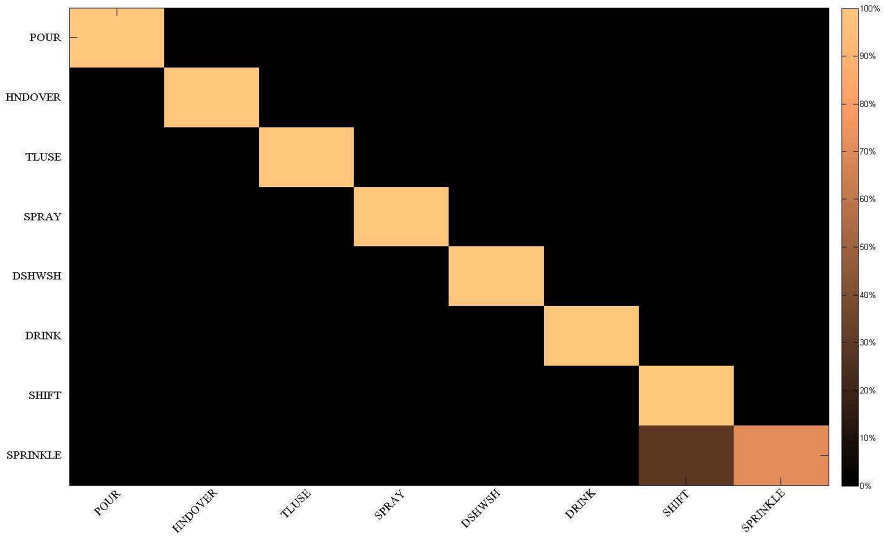


Figure 5.10: Confusion Matrix for ADLs inferred by HHMM/SVM Model at Level 1

5.6 Discussion

The HHMM/SVM hybrid model appears an overall stronger inference engine. However, this is somewhat misleading when put into the correct context, and the author advocate for the benefits that a HHMM model exhibits over a HHMM/SVM hybrid model when the appropriate criteria to model real-life, complex manipulation tasks are taken into consideration, as described below.

5.6.1 *Missing Data*

One of the challenges in dealing with real-time applications (such as ours) is dealing with missing data. Data can be missing or inexact due to various factors such as erroneous or faulty measurements by the instruments or sensors, and missing attributes from one or more sensor. The discriminative nature of the SVM classifier, makes it less capable of handling *missing data*. In contrast, the HHMM, being a generative model, is more capable of learning in the presence of missing values, and often performs better when training set sizes are small [Raina et al., 2004]. This is mainly due to the EM learning methodology which optimizes the model over the whole dimensionality, and thus models all the relationships between variables in a more equal manner [Le and Bengio, 2002].

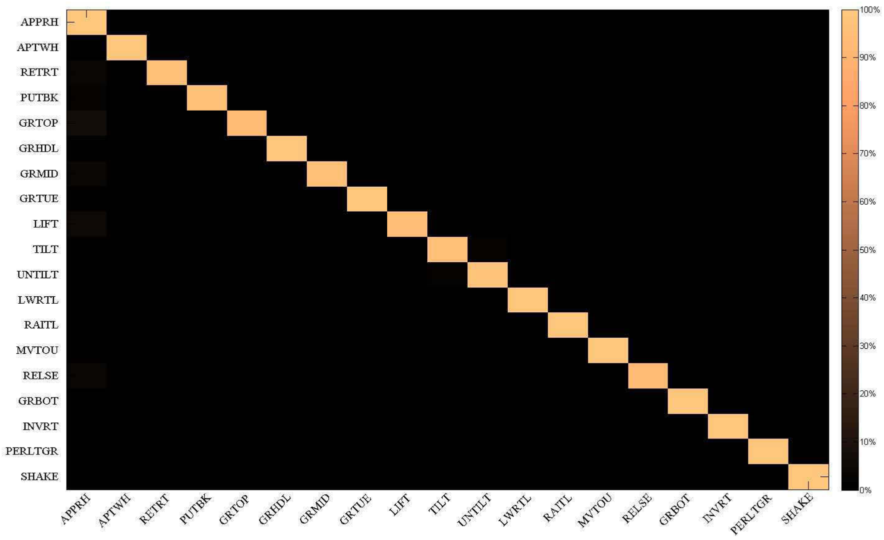


Figure 5.11: Confusion Matrix for APs inferred by HHMM/SVM Model at level 2

In order to emulate a case of missing data and smaller training data set, we conducted an experiment by randomly removing data samples from the training data. We divided the entire data set into two equal halves for training and testing (as we did for the HHMM experiments, specified in Section 5.5). The training data set was down-sized further by randomly sampling data at a frequency of *1/2 Hz, 1 Hz, 3 Hz, 5 Hz & 7 Hz*. By generating random data sets using this method, the information related to a given activity or AP which was lost due to down sampling can be regarded as representing missing or lost data. Note that the random sampling of data was done such that there would be at least one sample which represents an AP in any given activity sequence, so the down sample rates are approximate. This was done so as to maintain the representation of sequence of APs in any given activity. Further, to quantitatively analyse the impact of a smaller data set and of missing data on the performance of the HHMM and HHMM/SVM hybrid model, we generated 10 random training data sets for each case (i.e. 10 different data sets for *1/2 Hz, 1 Hz* etc.). Each of the trained models was then tested with a single testing data set which was sampled at *7 Hz*. Note that samples used for testing are separate, and do not overlap with any of the training data sets.

Figure 5.12 plots the mean and variance of the inference accuracy of the two models.

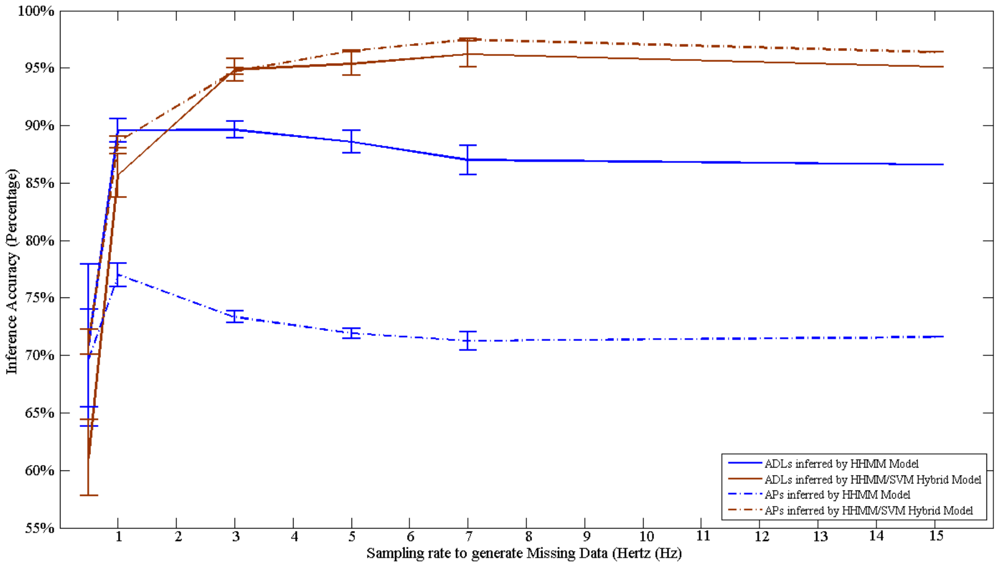


Figure 5.12: Comparison of inference accuracy of HHMM and HHMM/SVM Hybrid Model when training the model with varying amount of missing data

It can be seen how the performance of both models decreases substantially when the amount of missing data is around 97% of the full training data (at a sample rate of *1/2 Hz*). The inference accuracy of the HHMM/SVM hybrid model gradually increases as more training data becomes available. Conversely, the inference accuracy of the HHMM model remains almost constant despite the model being trained with varying amounts of training data. Hence the HHMM model seems better suited to generalise in the presence of missing data, as compared to the HHMM/SVM hybrid model.

5.6.2 *Testing with unseen Activity Sequence*

To further strengthen our advocacy of the HHMM model over HHMM/SVM hybrid models, we performed an experiment where we trained both models with 3 of the 4 sequences for each of the activities, and tested it with the unseen 4th sequence. This is different than the previous testing set used in Section 5.5.1 where data from all the 4 sequences sampled at *15 Hz* used for training. For this experiment we used data down sampled at *7 Hz*, as the experiment in Section 5.6.1 showed no measurable improvement at the higher rate. As can be seen in Figure 5.13, the HHMM model infers the long term activities with an accuracy of 74%, whereas the HHMM/SVM hybrid model's

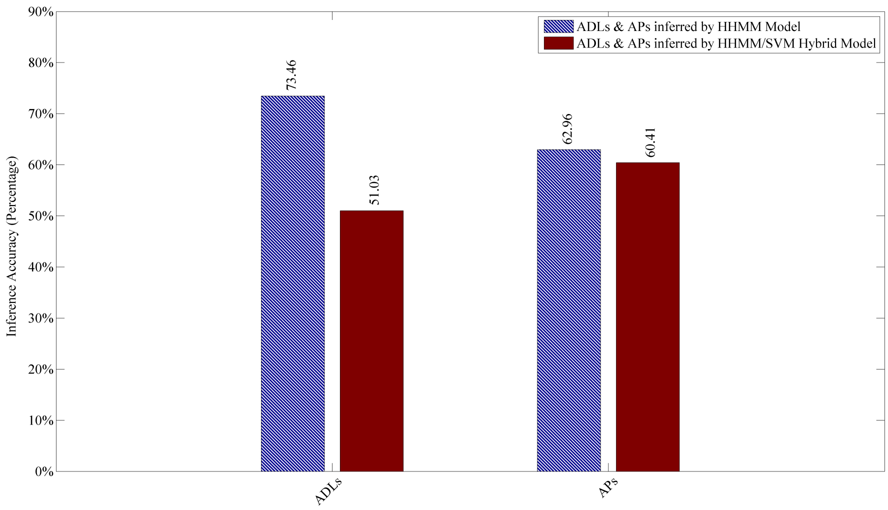


Figure 5.13: Activities and APs inferred by the HHMM and HHMM/SVM hybrid model when tested with unseen data

inference accuracy floats around 51%. Similarly APs were inferred with an accuracy of 63% by the HHMM model and 60% by HHMM/SVM hybrid model. The HHMM model outperforms the HHMM/SVM hybrid model in inferring both long term activities and APs, which further validates the better generalisation characteristics of the HHMM model.

5.6.3 Unsupervised Learning

Beyond the significant advantage of using HHMM models given their inherent generalization capabilities from smaller data sets, their unsupervised learning nature can not be under estimated. It significantly overcomes the rather difficult and costly process of obtaining labelled data for training. Moreover, unsupervised learning also opens the door for incorporating on-line learning algorithms whereby novelty in the patterns of performing an activity can be accomplished within the HHMM framework, e.g. using on-line-EM [Cappe and Moulines, 2009], a work currently under way. The modular nature of the HHMM model is therefore better equipped for real-time addition/deletion/modification in the state space [Dindo and Schillaci, 2010], a proposition which is less attractive when using generative models such as SVM where full re-training might be required.

5.7 Summary

In this chapter we presented a novel approach to infer users' manipulative activities using a HHMM based probabilistic model. The HHMM framework allows to flexibly divide an activity into a hierarchy, where complex ADLs are regarded as sequential combinations of more primitive building actions, or APs. The framework was tested on a set of manipulative sequences collected for different objects used in everyday life. The hierarchical framework proved to be a powerful tool to divide activities both vertically for natural language description of different activities from APs, and horizontally where continuous observations are clustered into different APs.

We also compared the inference accuracies of the HHMM model with a HHMM/SVM hybrid model, which performs learning in a semi-supervised manner and was in general able to infer more accurately at both APs and higher activity level. The model takes full advantage of the temporal characteristics of the HHMM model and the strong discriminating capability of the SVM classifier to infer APs and the related ADLs. However, it was shown to be less able to generalise in the absence of rich datasets, a well-known trade-off between generative and discriminative models. It is important to note that the inference was performed using off-line inference algorithm as compared to both on-line and off-line inference performed in the previous application (described in Chapter 3 and 4) where the ADLs were modelled for users of different assistive devices. The current application of modelling grasping and manipulation activities is more targeted towards robots learning ADLs from their human counterpart and once learned, perform those activities independently.

Chapter 6

Conclusion

6.1 Summary

In this thesis we introduced a novel approach to model and learn a wide variety of Activities of Daily Living (ADLs) as perceived by low level sensors. We deployed a strategy where complex ADLs are modelled by decomposing them into simpler atomic actions called Action Primitive (APs). The proposed approach has been motivated by evidence in biology and neuroscience which postulates that human motion behaviour is composed of simple, atomic movements that can be sequenced to form complex behaviour.

Further, we utilised a unified probabilistic framework capable of modelling both ADLs and the associated APs using low level sensor measurement. The ADLs that were modelled consists of locomotion, support, as well as grasping and manipulation activities that can be performed using the assistance of different robotic devices.

The proposed methodology of using a dictionary of APs proves to be an attractive approach for representing a broad spectrum of activities performed by humans. The primary advantage of representing ADLs using a string of APs relates to scalability whereby complex activities can be represented using a set of defined APs, and re-usability, where APs can be re-used in different sequences to construct any ADL. The Hierarchical Hidden Markov Model (HHMM) based probabilistic framework utilised to model ADLs and the corresponding APs proves to be an efficient tool capable of modelling both the uncertainty involved in human behaviour while performing different activities, and the noisy sensor data. The HHMM framework allows the ADLs to be flexibly divided into hierarchical representations, where ADLs are regarded as sequential combinations of more primitive APs.

As demonstrated in Chapter 3, we utilised a HHMM framework to model a wide variety of ADLs performed by power walker users. The pool of activities consist of navigational (visiting location of interest) and support activities (assistance to stand up). These behaviours were perceived through low level sensors fitted on the walker platform. As the navigational ADLs are long term in nature, they were modelled using a topological representation of an indoor environment, where location of interest were connected using junction points and segments. The junction points acted as intermediate navigational cues to model the overall ADLs. A pre-requisite for modelling ADLs using this approach was to generate a topological representation of the environment. In Chapter 4 we utilised human motion primitives to model ADLs, a further improvement over the previous approach. The use of a topological map was made redundant by modelling ADLs using intrinsic human motions. We modelled ADLs performed by users of two widely used mobility devices : a power walker and a robotic wheelchair. The dictionary of APs were developed by decomposing ADLs into atomic actions which were based on intrinsic human motion. The HHMM framework utilised in both methodologies (Chapter 3 and Chapter 4), modelled ADLs at the higher level by exploiting the temporal dependencies amongst APs, while the APs at the intermediate levels were modelled using human behaviour and the environment as perceived by the sensors. The inference for modelling ADLs in Chapter 3 and Chapter 4 was done using both inference techniques i.e. off-line inference algorithm in the form of maximum likelihood estimator and on-line inference using forward algorithm. The high inference accuracy obtained using both these techniques further strengthens our proposition of modelling ADLs using hierarchical probabilistic framework. The on-line inference algorithm which infers the ADLs and APs in real time offers the the ability for the control system to provide the necessary support to the user as and when required based on their behaviour.

In Chapter 5 we extended the applicability of the HHMM framework to model activities related to grasping and manipulation of everyday objects. We successfully utilised the ADL - AP representation to model these activities. The data which captured activities performed by humans consisted of features of hand-object data tracked in Cartesian space using a RGB-D kinect sensor. The ADLs were modelled using a HHMM framework by decomposing them into APs.

To draw a complete picture and thereby highlight the benefits of the proposed approach, we compared the inference accuracy of the HHMM framework with more traditional discriminative models (Support Vector Machines), other generative models (layered Dynamic Bayesian Networks) and combinations of both discriminative and generative models. On comparing the results, all the models predicted the APs with good accuracies. However, successful inference of complex ADLs was substantially reduced in the case of layered DBN and SVM models, validating the thesis proposal that the combination of decomposing ADLs at multiple levels and exploiting their inherent temporal nature plays a critical role in predicting complex interactive activities.

6.2 Future Research Direction

While the methodologies introduced in this thesis produces good results for what is known to be a difficult problem, this section will introduce potential improvements and propose applications that can be targeted in future work. Section 6.2.1 discusses the real-time implementation and testing of the framework on robots. Section 6.2.2 discusses possible modelling improvements that could be added to the HHMM framework to model ADLs.

6.2.1 Testing the Framework to Control Robots

Efficient control of the robots based on the desire/intention of the user becomes an important criterion for a tightly knit human-robot interactive system. The HHMM framework used in this work has been tested off-line with real-time data collected while user performed different ADLs using the support of the walker and/or wheelchair. In order to control the robots (particularly the walker) as per the user's intention, a pool of APs was generated such that it could directly be mapped to the control system of the robot. We tested our proposition by mapping some of the user intentions inferred by the Dynamic Bayesian Network [Patel et al., 2010] to the control policy of the walker. The proposed control policy can be extended to the APs inferred by the HHMM framework to control the robot as per the user's interaction. Similarly, the APs defined for a robotic manipulator to learn grasping and manipulation activities from a human teacher are such that they can be transferred to the control policy of the robotic arm. Despite the difference between the kinematic model and DOFs of the human arm and the robotic manipulator, the interactions between the robotic arm and the objects in their surrounding (e.g. grasping the object with a particular pose in order to accomplish the desired activity) will be of a similar nature to their human teacher.

6.2.2 Enhancing the Probabilistic Framework

Adaptive Learning

The HHMM framework introduced for modelling ADLs has not been exploited to its full potential in this thesis. For instance the number of ADLs and APs inferred by the HHMM framework is fixed, which limits the ability of the model to adapt to change and/or addition of new ADLs/APs. The technique described in [Dindo and Schillaci, 2010], can be utilised so as to make the framework more adaptive to accommodate the change in user activities.

On-line Learning

Learning is another area where the capability of the model can be exploited further. In this thesis we deployed an off-line EM technique which learns and optimises the parameters of the model from the data. The online-EM learning technique proposed in [Cappe and Moulines, 2009] can be applied to learn the parameters of the model in real-time. The learning starts with the rough model of the ADLs/APs and is further optimised as more data related to user behaviour becomes available.

Structure Evolution/Learning

In this thesis, the structure of the model which defines the dependencies between the parameters (i.e. observations and user states) is manually defined. With the increase in complexity of user states and data (e.g. the data used in Chapter 5), defining the structure manually does not exploit the true relationship between different parameters. Structure learning techniques for Bayesian Network [Eaton, 2007], [Francois, 2004], [Chickering, 2002] can be utilised, which explores feature dependencies to model user behaviours.

Structure learning approach can also be extended to learn the temporal dependencies in case of Dynamic Bayesian Network [Lähdesmäki and Shmulevich, 2008] and Hidden Markov Model (HMM) [Kulic and Nakamura, 2010] which exploits both static dependencies between states and sensors and temporal state dependencies over time. Similarly, to explore hierarchical dependencies present in data, non-parametric bayesian methods described in [Wang et al., 2007] can be utilised, which autonomously discovers the optimal number of levels in the hierarchy as well as the states and parameter at each level.

6.2.3 Automatic Generation of Action Primitive (APs) structure

The primary aim of this thesis was to model ADLs using a dictionary of APs. The APs were generated based on visual inspection and intuition of how ADLs were performed. The subsequent step from this thesis that can further be explored would be to automatically generate the pool of APs based on the ADLs. Techniques described in [Lee et al., 2013] can be utilised to generate APs which are frequently encountered in ADLs.

6.3 Conclusion

This thesis contributed to improving, extending and generalising HRI particularly in the field of assistive robotics. It illustrated how the approach of modelling activities by decomposing then into a string of atomic actions can be utilised to model a variety of high level activities from low level sensor measurements. Different human assistance domains were targeted successfully, and a general interaction framework based on Hierarchical HMM was proposed. Robotic systems will have to acquire more sophisticated assistive capabilities if they are to operate in unstructured, dynamic, human-centred environments, responsive to the needs of interacting with humans. In that context, awareness of human intentions play a key role in being able to apply any practical assistive action where a user interacts with a robot to carry out their regular daily activities, be that navigational, grasping and manipulation, communicative, or others. The methodologies and strategies introduced in this thesis can be extended to other application domains in order to achieve a natural, sophisticated interaction synergy between humans and robots.

Appendix A

Power Walker

The power walker shown in Figure A.1 was used as an aid device to perform various ADLs. It is a modified commercial rollator walking frame with four wheels. The base frame had been instrumented with '$24Volts$' (DC) reversible gear-head motors, rotary mechanical couplings and incremental optical encoders to the two rear wheels (Figure A.2(b)) (front wheels are passive). The motors were PWM driven using a national semiconductor LMD18200 3A, 55V H-Bridge motor driver. They were specifically deployed to provide active ambulatory support to the user.

The user's behaviour was perceived by the walker through a set of low level sensors installed on the walker. Four strain gauges (SGs)(two on each handle bar) (Figure A.2(a)) were used to measure the pressure a user would be exerting while handling the walker. The differential forces between the vertical axes in each handle-bar are indicative of the users' readiness to start a task (sitting down, standing up or ambulation steering). The strain gauges were micro measurements 120($ohms$) and a full Wheatstone bridge electronics circuit was used to measure the pressure exerted by the user on each individual handle of the walker.

Apart from strain gauges there were two proximity sensors (Sharp GP2Y0A02YK) (fitted in front of the walking frame as can be seen in Figure A.1(b)) which were used to sense behaviours such as the proximity of the user from the walker, user present

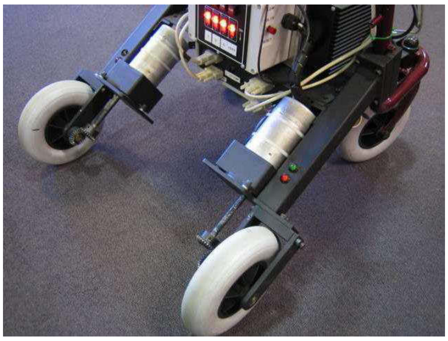
(a) Front view of rollator Power Walker

(b) Rear view of rollator Power Walker

Figure A.1: Front and rear view of the power walker

in front of the walker, user in sitting or standing position etc. The sensing range of the proximity sensors after calibration was between [20, 150] cm. The hardware also included a radio switch, which indicates to the walker to come back from its parking position to where it last left the user, or vice-versa. This feature is more advantageous in certain locations such as the living room, or bedroom where the user spends more time unaided. The platform was also equipped with a Hokuyo URG-04LX laser range finder at the front, which was utilised for localising the walker in a given environment and to actively safeguard from static and dynamic obstacles present in an environment.

(a) Strain Gauge



(b) wheel encoder and motors

Figure A.2: (a) Strain gauge installed on the handles of the walker and (b) Wheel encoders and DC motors installed on the power walker

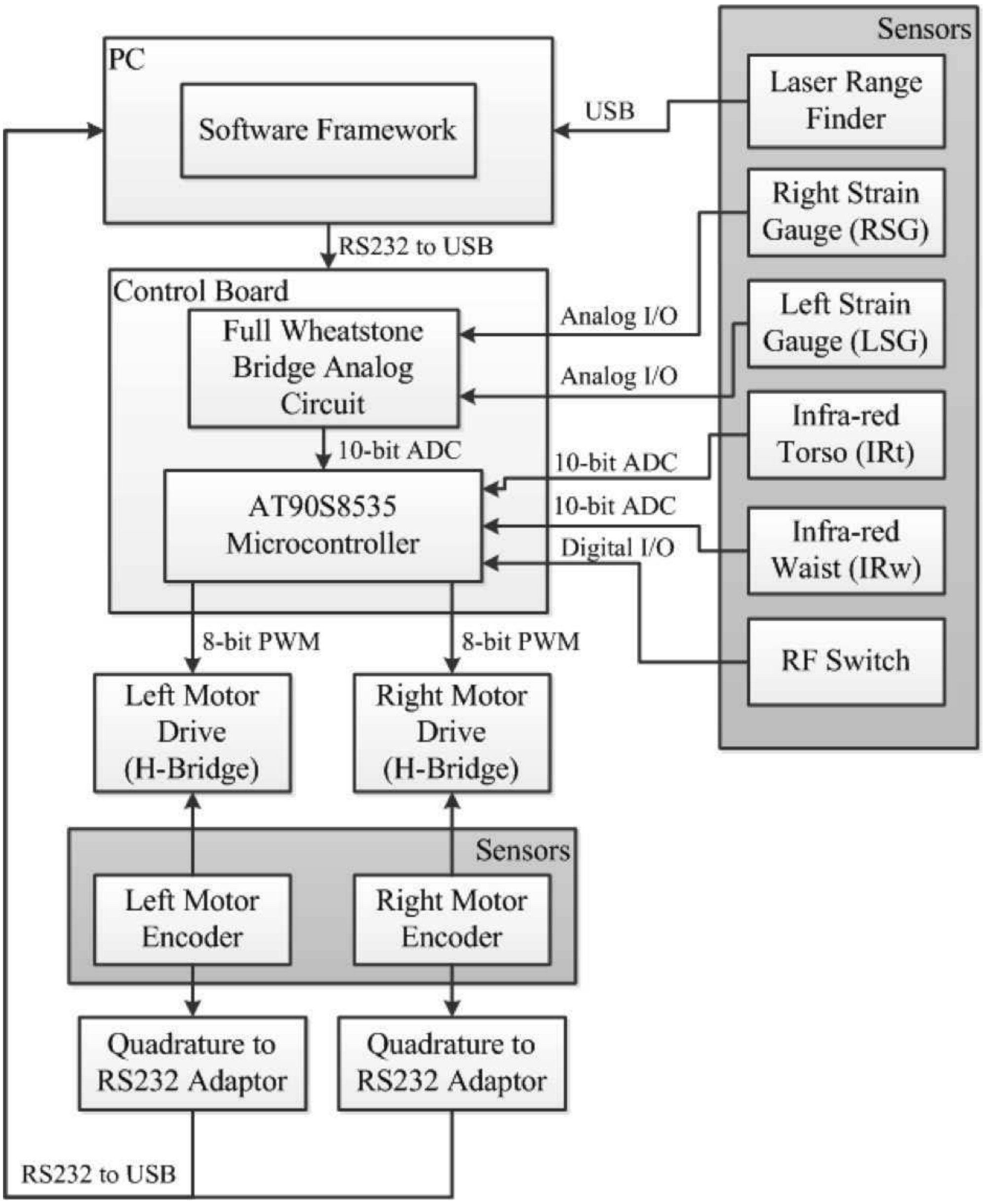


Figure A.3: Electronics system integration on the power walker platform

In addition to the sensing hardware the walker was also instrumented with a low-level micro-controller which communicates with all the sensors and controls the speed of the walker. The platform was also equipped with a high-level computer for data processing and storage. The low-level micro-controller acted as a communication bridge between the sensors and the high level computer for data processing. A detailed system diagram depicting the communication between the various sensors, micro-controller and high

Users	User 1		User 2		User 3	
Gait Parameters	Normal Walking	Speed Limit Walking	Normal Walking	Speed Limit Walking	Normal Walking	Speed Limit Walking
Stance Phase (sec)	0.46 (39.95%)	0.96 (80.26%)	0.58 (41.20%)	1.17 (81.59%)	1.30 (70.29%)	1.63 (82.28%)
Swing Phase (sec)	0.78 (60.05%)	0.24 (19.73%)	0.74 (58.79%)	0.28 (18.40%)	0.33 (29.71%)	0.35 (17.71%)
Avg. Speed (m/sec)	0.52	0.16	0.55	0.19	0.38	0.25

Table A.1: Temporal-Distance Gait Parameters for each user

level computer is shown in Figure A.3.

A.1 Gait Characteristics

As the author did not have access to aged or frail population to perform the experiments needed to fully validate this work, data were collected using healthy volunteers. As per the results presented by Zong *et.al.* [Zong et al., 2010], the basic gait patterns such as *Gait cycle, Stance phase, Swing phase* and *Walking speed* of healthy users are different to that of frail individuals. Hence in order to obtain representative data having gait characteristics similar to those of elderly or frail individuals, we collected data from 3 healthy users' (1 male, 2 female) by controlling the maximum speed of the power walker.

To further validate our approach, we collected data whereby the user was asked to walk twice on a straight, flat surface path for $10m$ using the power walker. In the first part, users were asked to walk at their normal everyday walking pace. In the second part, users were asked to walk the same distance and path with the speed of the walker controlled such that the maximum speed at which user could walk was set at $0.3m/sec$. The gait characteristics achieved using this speed were similar to that of a typical frail and elderly user of a mobility assistance platform as reported in the literature [Zong et al., 2010]. The only sensor data used for this experiment were obtained from the IR sensor installed beneath the walker, which records the gait dynamics of the user, a setup similar to that used by Zong *et. al.* [Zong et al., 2010].

Results

By setting an appropriate maximum walker speed, user's gait characteristics (Table A.1) were found to be in agreement with those reported in the literature [Zong et al., 2010]. The temporal-distance gait parameters reflect the persons' dynamics during walking. Gait parameters as reported in [Zong et al., 2010] were extracted to analyse the gait patterns of the user for the data collected as described above. The user spent more time in the stance phase as compared to the swing phase, and the walking speed was also reduced when compared to their normal walking gait parameters. The results of this experiment substantiate the fact that by controlling the maximum speed of the walker, the data logged for the ADLs inference experiments from healthy subjects can be assumed to closely correlate with the gait characteristics of an old/frail person. Furthermore, the variation in the user's gait dynamics when controlling the speed of the walker, as compared with their normal walking pattern, was also found to be in close correlation to that reported in [Zong et al., 2010].

Appendix B

Robotic Wheelchair

The wheelchair used for experimentation, depicted in Figure B.1, is a commercially available power wheelchair (Invacare rollar M1 [Invacare, reviewd on 3rd January 2013]) modified with the necessary hardware. It was instrumented with a computer (attached behind the backrest), wheel encoders, and a Hokuyo URG-04LX laser range finder used for localisation. It also had two differentially driven wheels at the rear, and two passive casters at the front. It measures 1.2 x 0.7m, certainly a large robot when driving around in a typical office environment with narrow passages, long corridors, and cluttered static obstacles. The wheelchair can travel at a maximum speed of up to 15km/h. The Hokuyo URG- 04LX laser range finder is located on a special stand at the front of the wheelchair. The wheelchair is further instrumented with a general Input/Output (I/O) board (capable of integrating both digital and analog signals) and a specialised circuitry called the Wheelchair Interface Unit (WIU). With the instrumented electronics the wheelchair is capable of operating in two modes: manual and autonomous. In manual mode the wheelchair can be controlled using the joystick whereas in autonomous mode the wheelchair is controlled by the computer. The WIU is specifically designed so as to control the wheelchair in autonomous mode as it converts the signal received from the computer to standard command signals that control the normal functioning of the wheelchair as if a user was controlling it, such as activating the motors, increasing/decreasing the gears, or sounding the horn. Similar to the walker hardware, the wheelchair was also instrumented with a radio switch, which indicates to the wheelchair to come back from its parking position to where it last left the user, or vice-versa. Details of hardware and communication between different sensors and he motor controller are shown in Figure B.2.

(a) Front view of Robotic Wheelchair (b) Rear view of Robotic Wheelchair

Figure B.1: Front and rear view of the robotic wheelchair

The support provided by a wheelchair used in this work is mainly navigational in nature. The user behaviour to perform a specific ADL is perceived through the joystick of the wheelchair, where the user provides the navigational cues of where/which direction the he/she intends to go.

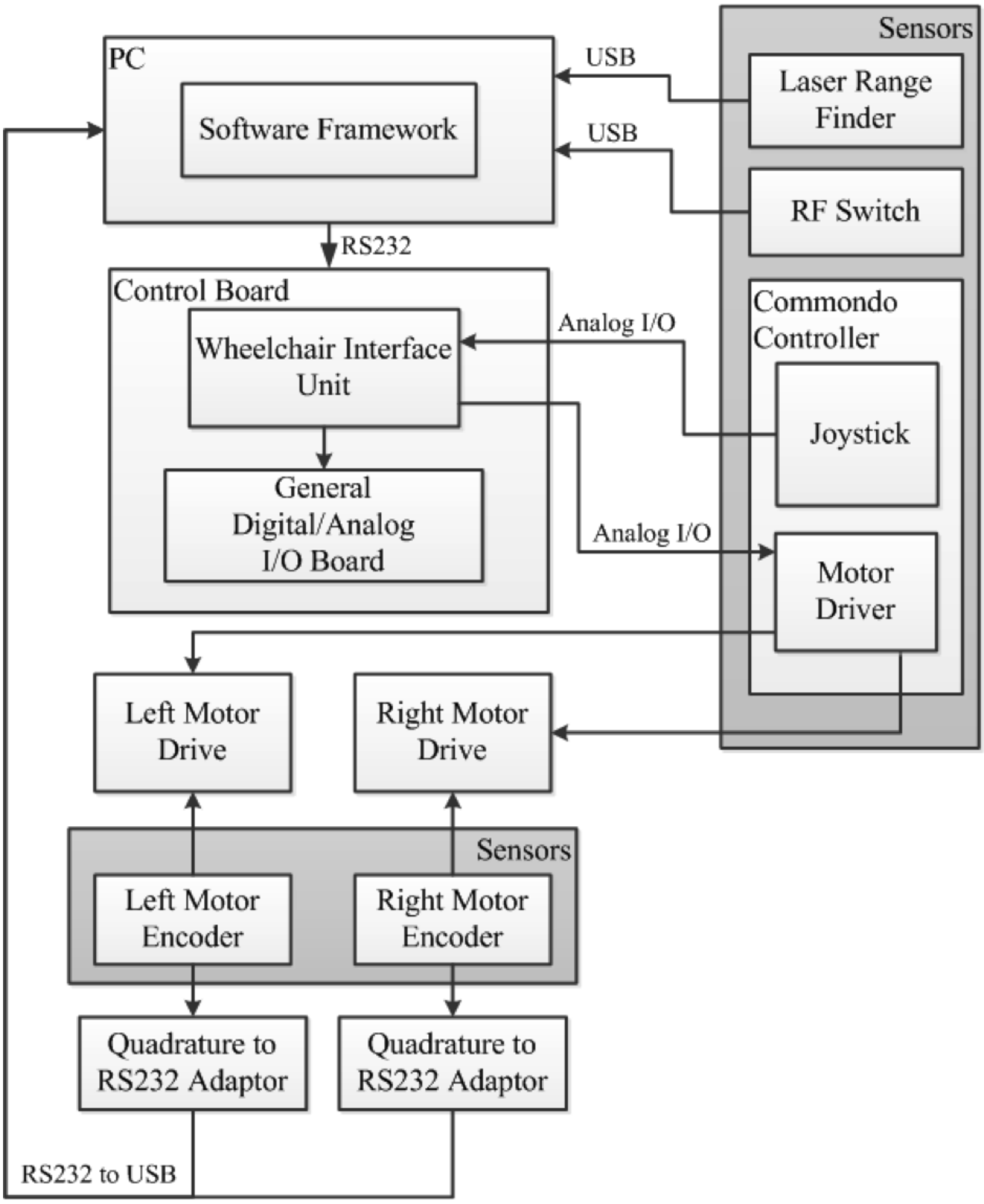


Figure B.2: Electronics system integration on the robotic wheelchair platform

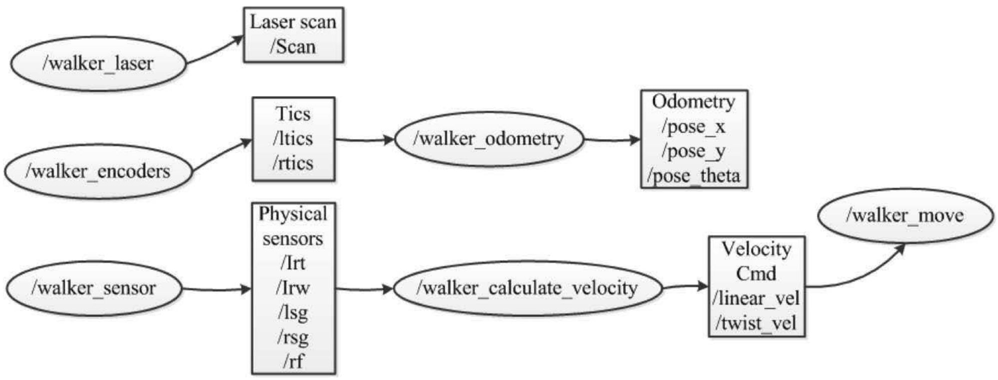
Appendix C

Software Framework

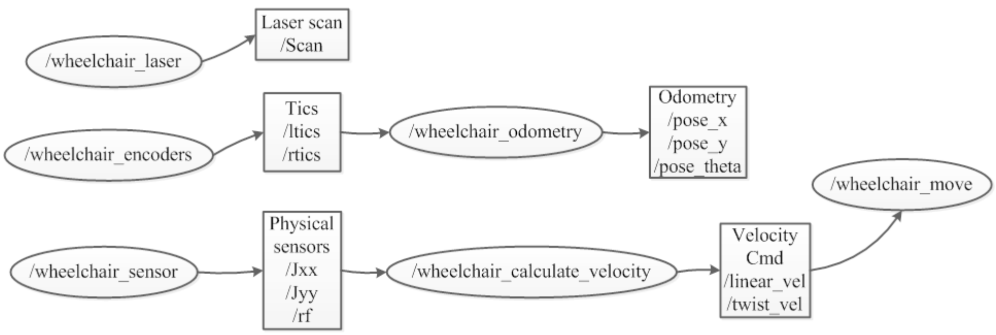
Software drivers used to communicate with the low level sensors and to send commands to control the actuators of both the mobility devices were written using the framework of Player/Stage [Gerkey et al., 2003] and later on migrated to Robotics Operating System (ROS) [Quigley et al., 2009]. The framework ensured integration and re-usability of the sensors present on both the robotic platforms with various software drivers such as Adaptive Monte-Carlo Localisation (AMCL), different path-planning and obstacle avoidance algorithms readily available within these frameworks.

Player/Stage provides a network interface to a variety of robot and sensor hardware. Player's client/server model allows robot control programs to be written in any programming language and to run on any computer with a network connection to the robot. Player supports multiple concurrent client connections to devices, creating new possibilities for distributed and collaborative sensing and control. The player proxy driver capable of publishing sensor information which can further be fused with other sensors was developed in house reported in [Osswald, 2008], which were compatible with Player version 2.2. Similarly, the player proxy driver for physical sensor of wheelchair was compatible with Player version 3.0. With further development of ROS and its wider acceptability and support within the robotics community we migrated the drivers for both the platforms from Player/Stage to ROS. ROS uses a distributed peer-to-peer node architecture which allows each functional block to operate as a completely independent node. The drivers for both the platforms were developed by graduate students at the University of Technology Sydney, reported in [Kelaher et al., 2012].

For all our experiments we collected data using the Player/Stage client-server architecture, which allowed us to save all the sensor information such as localisation coordinates $(x, y, theta)$, physical sensors on the platforms along with time stamps. Figure C.1 depicts the sensor and the type of information published by each sensor which is further fused with other sensor data to derive different information such as location of the robot (walker/wheelchair), path planning and obstacle avoidance.



(a) Data published by each sensor on power walker



(b) Data published by each sensor on robotics wheelchair

Figure C.1: Sensor and type of data published by each player proxy nodes which is further utilised/fused with other sensor information

Bibliography

M. Alwan, G. Wasson, P. Sheth, A. Ledoux, and C. Huang. Passive derivation of basic walker-assisted gait characteristics from measured forces and moments. In *IEEE 26th Annual International Conference of Engineering in Medicine and Biology*, pages 2691 – 2694, 2005. [pp.]

A. Atrash and J. Pineau. A bayesian reinforcement learning approach for customizing human-robot interfaces. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, pages 355 – 360, 2009. [pp.]

F. Aurenhammer. Voronoi diagrams-a survey of a fundamental geometric data structure. *ACM Computing Survey*, 23(3):345 – 405, 1991. [pp.]

C. M. Bishop and J. Lasserre. Generative or discriminative? getting the best of both worlds. In *Bayesian Statistics 8*, pages 3 – 24. Oxford University Press, 2007. [pp.]

J. Blimes. A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical Report ICSI-TR-97-021, University of Berkeley, 1998. [pp.]

S. W. Brose, D. J. Weber, B. A. Salatin, G. G. Grindle, H Wang, J. J. Vazquez, and R. A. Cooper. The role of assistive robotics in the lives of persons with disability. *American Journal of Physical Medicine and Rehabilitation*, 89(6):509 – 521, 2010. [pp.]

M. Buettner, R. Prasad, M. Philipose, and D. Wetherall. Recognizing daily activities with rfid-based sensors. In *The 11th International Conference on Ubiquitous Computing*, pages 51 – 60, 2009. [pp.]

H. H. Bui, S. Venkatesh, and G. West. Policy recognition in the abstract hidden markov model. *Journal of Artificial Intelligence Research*, 17:451 – 499, 2002. [pp.]

O. Cappe and E. Moulines. Online expectationmaximisation algorithm for latent variable models. *Journal of Royal Statistical Society: Series B Statistical Methodology*, 71(3):593–613, 2009. [pp.]

S. Carberry. Techniques for plan recognition. *User Modeling and User-Adapted Interaction*, 2001. [pp.]

T. Carlson and Y. Demiris. Increasing robotic wheelchair safety with collaborative control: Evidence from secondary task experiments. In *IEEE International Conference on Robotics and Automation*, pages 5582 – 5587, 2010. [pp.]

A. Castellani, D. Botturi, M. Bicego, and P. Fiorini. Hybrid hmm/svm model for the analysis and segmentation of teleoperation tasks. In *IEEE International Conference on Robotics and Automation, 2004*, volume 3, pages 2918 – 2923, 2004. [pp.]

E. Castillo, J. M. Gutierrez, and A. S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer Verlag, 1997. [pp.]

D. M. Chickering. The winmine toolkit. Technical Report MSR-TR-2002-103, Microsoft Research, 2002. [pp.]

Congressional Budget Office. Global population aging in the 21st century and its economic implications, 2005. [pp.]

E. Demeester, A. Huntemann, D. Vanhooydonck, G. Vanacker, A. Degeest, H. Van

Brussel, and M. Nuttin. Bayesian estimation of wheelchair driver intents: Modeling intents as geometric paths tracked by the driver. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5775 – 5780, 2006. [pp.]

E. Demeester, A. Hüntemann, D. Vanhooydonck, G. Vanacker, Van B. H., and M. Nuttin. User-adapted plan recognition and user-adapted shared control: A bayesian approach to semi-autonomous wheelchair driving. *Autonomous Robots*, 24(2):193 – 211, 2008. [pp.]

H. Dindo and G. Schillaci. An adaptive probabilistic approach to goal-level imitation learning. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4452 – 4457, 2010. [pp.]

S. Dubowsky, F. Génot, S. Godding, H. Kozono, A. Skwersky, H. Yu, and L. S. Yu. PAMM - A robotic aid to the elderly for mobility assistance and monitoring: A helping-hand for the elderly. In *IEEE International Conference on Robotics and Automation*, pages 570 – 576, 2000. [pp.]

D. Eaton. Bayesian structure learning using dynamic programming and mcmc. In *Uncertainty in Artificial Intelligence*, pages 101 – 108, 2007. [pp.]

L. Fehr, W. E. Langbein, and S. B. Skaar. Adequacy of power wheelchair control interfaces for persons with severe disabilities: a clinical survey. *Journal of rehabilitation research and development*, 37(3):353 – 360, 2000. [pp.]

S. Fine, Y. Singer, and N. Tishby. The hierarchical hidden markov model: Analysis and applications. *Machine Learning*, 32:41 – 62, 1998. [pp.]

O. Francois. Bnt structure learning package: documentation and experiments. Technical report, Laboratoire PSI, Universitè et INSA de Rouen, 2004. [pp.]

L. Gerard and D. M. Kenneth. The application of robotics to a mobility aid for the elderly blind. *Robotics and Autonomous Systems*, 23(4):245 – 252, 1998. [pp.]

B. Gerkey, R. Vaughan, and A. Howard. The player/stage project: Tools for multi-robot and distributed sensor systems. In *The 11th IEEE International Conference on Advanced Robotics*, pages 317 – 323, 2003. [pp.]

H. Ghasemzadeh, J. Barnes, E. Guenterberg, and R. Jafari. A phonological expression for physical movement monitoring in body sensor networks. In *IEEE International Conference on Mobile Ad Hoc and Sensor Systems*, pages 58 – 68, 2008. [pp.]

J. Glover, S. Thrun, and J. T. Matthews. Learning user models of mobility-related activities through instrumented walking aids. In *IEEE International Conference on Robotics and Automation*, pages 3306 – 3312, 2004. [pp.]

C. Heinze. *Modeling Intention Recognition for Intelligent Agent Systems*. PhD thesis, The University of Melbourne, 2003. [pp.]

Y. Hirata, A. Muraki, and K. Kosuge. Motion control of intelligent passive-type walker for fall-prevention function based on estimation of user state. In *IEEE International Conference on Robotics and Automation*, pages 3498 – 3503, 2006. [pp.]

J. Hoey, A. V. Bertoldi, P. Poupart, and A. Mihailidis. Assisitng persons with dementia during handwashing using a partially observable markov decision process. In *the International Conference on Vision Systems*, March 2007. [pp.]

S. Iba, C. J. J. Predis, and P. K. Khosla. Intention aware interactive multi-modal robot programming. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 3479 – 3484, 2003. [pp.]

A.J. Ijspeert, J. Nakanishi, and S. Schaal. Movement imitation with nonlinear dynamical systems in humanoid robots. In *IEEE International Conference on Robotics and Automation*, volume 2, pages 1398–1403, 2002. [pp.]

Invacare. Invacare rollar m1 parts manual, reviewd on 3rd January 2013. URL http://www.invacare.com.au/index.cfm/1,135,584,40,html/Roller-M1. [pp.]

F. V. Jensen. *An Introduciton to Bayesian Networks*. UCL Press, 1996. [pp.]

Kyung-Nam K. and R.S. Ramakrishna. Vision-based eye-gaze tracking for human computer interface. In *IEEE International Conference on Systems, Man, and Cybernetics*, volume 2, pages 324 – 329, 1999. [pp.]

T. Kanda, D.F. Glas, M. Shiomi, and N. Hagita. Abstracting people's trajectories for social robots to proactively approach customers. *IEEE Transactions on Robotics*, 25 (6):1382 – 1396, 2009. [pp.]

S. Kang, Y. Ho, and I. Hyuk Moon. Development of an intelligent guide-stick for the blind. In *IEEE International Conference on Robotics and Automation*, volume 4, pages 3208 – 3213, 2001. [pp.]

D. Kawanaka, T. Okatani, and K. Deguchi. Hierarchical-hmm based recognition of human activity. *Proc of Machine Vision Applications*, 2005. [pp.]

N. Kelaher, C. Kerr, and H. Lowes. Ros driver for cas walker (technical report), 2012. [pp.]

S.M. Khansari-Zadeh and A. Billard. Imitation learning of globally stable non-linear point-to-point robot motions using nonlinear programming. In *IEEE/RSJ International conference on Intelligent Robots and Systems*, pages 2676 – 2683, 2010. [pp.]

B. Kluge, C. Kohler, and E. Prassler. Fast and robust tracking of multiple moving objects with a laser range finder. In *IEEE International Conference on Robotics and Automation*, volume 2, pages 1683 – 1688, 2001. [pp.]

P. Krauthausen and U.D. Hanebeck. Intention recognition for partial-order plans using dynamic bayesian networks. In *12th International Conference on Information Fusion*, pages 444 – 451, 2009. [pp.]

V. Krüger, D. Herzog, S. Baby, A. Ude, and D. Kragic. Learning actions from observations. *IEEE Robotics & Automatation Magazine*, 17(2):30 – 43, 2010. [pp.]

D. Kulic and Y. Nakamura. Incremental learning of human behaviors using hierarchical hidden markov models. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4649 – 4655, 2010. [pp.]

D. Kulic, D. Kragic, and V. Krüger. Learning action primitives. In *Visual Analysis of Humans*, pages 333 – 353. 2011. [pp.]

H. Lähdesmäki and I. Shmulevich. Learning the structure of dynamic bayesian networks from time series and steady state measurements. *Machine Learning*, 71(2-3):185 – 217, 2008. [pp.]

Q. Le and S. Bengio. Hybrid generative-discriminative models for speech and speaker recognition. Idiap-RR Idiap-RR-06-2002, IDIAP, 2002. [pp.]

K. Lee, Y. Su, T. Kim, and Y. Demiris. A syntactic approach to robot imitation learning using probabilistic activity grammars. *Robotics and Autonomous Systems*, 61(12):1323 – 1334, 2013. [pp.]

L. Liao, D. Fox, J. Hightower, H. Kautz, and D. Schulz. Voronoi tracking: location estimation using sparse and noisy sensor data. In *IEEE/RSJ Conference on Intelligent Robots and Systems*, volume 1, pages 723 – 728, 2003. [pp.]

L. Liao, D. J. Patterson, D. Fox, and H. Kautz. Learning and inferring transportation routines. *Artificial Intelligence*, 171:311 – 331, 2007. [pp.]

M. Luber, J.A. Stork, G.D. Tipaldi, and K.O. Arras. People tracking with human motion predictions from social forces. In *IEEE International Conference on Robotics and Automation*, pages 464 – 469, 2010. [pp.]

S. Mahadevan, M. Ghavamzadeh, K. Rohanimanesh, and G. Theocharous. *Hierarchical Approaches to Concurrency, Multiagency, and Partial Observability*, chapter 11, pages 285 – 310. Wiley-IEEE Press, 2004. [pp.]

C. Mandel, K. Huebner, and T. Vierhuff. Towards an autonomous wheelchair: Cognitive aspects in service robotics. In *Proceedings of Towards Autonomous Robotic Systems*, pages 165 – 172, 2005. [pp.]

A. C. Morris, R. R. Donamukkala, A. Kapuria, A. Steinfeld, J. Matthews, J. Dunbar-Jacobs, and S. Thrun. A robotic walker that provides guidance. In *IEEE Conference on Robotics and Automation*, volume 1, pages 25 – 30, 2003. [pp.]

K. P. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of Califronia, Berkeley, 2002. [pp.]

K. P. Murphy. *Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning series)*. The MIT Press, 2012. [pp.]

B. Nemec and A. Ude. Action sequencing using dynamic movement primitives. *Robotica*, 30(5):837 – 846, 2012. [pp.]

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 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 955 – 960, 2005. [pp.]

A. Officer and A. Posarac. *World Report on Disability*. World Health Organisation, 2011. [pp.]

I. Oikonomidis, N. Kyriazis, and A. Argyros. Full dof tracking of a hand interacting with an object by modeling occlusions and physical constraints. In *IEEE International Conference on Computer Vision*, pages 2088 – 2095, 2011a. [pp.]

I. Oikonomidis, N. Kyriazis, and A. Argyros. Efficient model-based 3d tracking of hand articulations using kinect. In *Proceedings of the British Machine Vision Conference*, pages 101.1 – 101.11. BMVA Press, 2011b. [pp.]

F. Omar, M. Sinn, J. Truszkowski, P. Poupart, J. Tung, and A. Caine. Comparative analysis of probabilistic models for activity recognition with an instrumented walker. In *the 26th Conference on Uncertainty in Artificial Intelligence*, pages 1 – 9, 2010. [pp.]

S. Osentoski, V. Manfredi, and S. Mahadevan. Learning hierarchical models of activity. In *IEEE/RSJ International Conference on Robots and Systems*, volume 1, pages 891 – 896, 2004. [pp.]

V. Osswald. Human-in-loop control of an automated walking aid platform (technical report), 2008. [pp.]

C. Papazov and D. Burschka. An efficient ransac for 3d object recognition in noisy and occluded scenes. *Lecture Notes in Computer Vision*, 6492:135 – 148, 2011. [pp.]

P. Pastor, H. Hoffmann, T. Asfour, and S. Schaal. Learning and generalization of motor skills by learning from demonstration. In *IEEE/RSJ International Conference on Robotics and Automation*, pages 1293 – 1298, 2009. [pp.]

M. Patel, J. V. Miró, and G. Dissanayake. Dynamic bayesian networks for learning interactions between assistive robotic walker and human users. In *The 33rd Annual German Conference on Artificial Intelligence*, pages 333 – 340, 2010. [pp.]

M. Patel, J. V. Miró, and G. Dissanayake. A hierarchical hidden markov model to support activities of daily living with an assistive robotic walker. In *4th IEEE/RAS EMBS International Conference on Biomedical Robotics and Biomechatronics*, pages 1071 – 1076, 2012. [pp.]

Occupational Therapy Practice. Occupational therapy practice framework : Domain & process (2nd edition). *The American Journal of Occupational Therapy*, 62(6): 625–683, 2008. [pp.]

A. Pruski, M. Ennaji, and Y. Morere. VAHM: A User Adapted Intelligent Wheelchair. *Proceedings of the International Conference on Control Applications*, 2:784 – 789, 2002. [pp.]

D. Pynadath. *Probablistic Grammars for Plan Recognition*. PhD thesis, The University of Michigan MI, 1999. [pp.]

M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng. Ros: An open-source robot operating system. In *ICRA Workshop on Open Source Software*, 2009. [pp.]

R. Raina, Y. Shen, A. Y. Ng, and A. McCallum. Classification with hybrid generative/discriminative models. In *Advances in Neural Information Processing Systems*. MIT Press, 2004. [pp.]

M. Raptis, K. Wnuk, and S. Soatto. Flexible Dictionaries for Action Classification. In *The 1st International Workshop on Machine Learning for Vision-based Motion Analysis*, 2008. [pp.]

D. Rawlinson and R. Jarvis. Ways to tell robots where to go - directing autonomous robots using topological instructions. *IEEE Robotics & Automation Magazine*, 15(2): 27 – 36, 2008. [pp.]

G. Rizzolatti, L. Foggassi, and V. Gallese. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2:661 – 670, 2001. [pp.]

A. Saxena, J. Driemeyer, and A. Y. Ng. Robotic grasping of novel objects using vision. *International Journal Robotics Research*, 27(2):157–173, 2008. [pp.]

S. Schaal, A. J. Ijspeert, and A. Billard. Computational approaches to motor learning by imitation. *Philosophical transaction of the Royal Society of London, series B*, 358 (1431):537 – 547, 2003. [pp.]

S. Schaal, J. Peters, J. Nakanishi, and A. Ijspeert. Learning movement primitives. In *International Symposium on Robotics Research*. Springer, 2004. [pp.]

O. C. Schrempf and U. D. Hanebeck. A generic model for estimating user intentions in human-robot cooperation. In *Proceedings of the 2nd International Conference on Informatics in Control, Automation and Robotics*, volume 3, pages 251 – 256, 2005. [pp.]

V. Setalaphruk, A. Ueno, I. Kume, Y. Kono, and M. Kidode. Robot navigation in corridor environments using a sketch floor map. In *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, volume 2, pages 552 – 557, 2003. [pp.]

D. Song, K. Huebner, V. Kyrki, and D. Kragic. Learning task constraints for robot grasping using graphical models. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1579 – 1585, 2010. [pp.]

D. Song, C. H. Ek, K. Huebner, and D. Kragic. Multivariate discretization for bayesian network structure learning in robot grasping. In *IEEE International Conference on Robotics and Automation*, pages 1944 – 1950, 2011a. [pp.]

D. Song, C. H. Ek, K. Huebner, and D. Kragic. Embodiment-specific representation of robot grasping using graphical models and latent-space discretization. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 980 – 986, 2011b. [pp.]

S. Srinivasa, D. Ferguson, M. Vandeweghe, R. Diankov, D. Berenson, C. Helfrich, and H. Strasdat. The robotic busboy: Steps towards developing a mobile robotic home assistant. In *International Conference on Intelligent Autonomous Systems*, pages 129–136, 2008. [pp.]

J. Stadermann and G. Rigoll. A hybrid svm/hmm acoustic modeling approach to automatic speech recognition. In *INTERSPEECH*. ISCA, 2004. [pp.]

T. Stiefmeier, D. Roggen, and G. Tröster. Gestures are strings: efficient online gesture spotting and classification using string matching. In *the ICST(Institute of Computer Sciences, Social-Informatics and Telecommunications Engineering) 2nd International Conference on Body Area Networks*, pages 16:1 – 16:8, 2007. [pp.]

T. Taha, J. V. Miró, and G. Dissanayake. Pomdp-based long-term user intention prediction for wheelchair navigation. In *IEEE International Conference on Robotics and Automation*, pages 3920 – 3925, 2008. [pp.]

K. A. Tahboub. Intelligent human-machine interaction based on dynamic bayesian networks probabilistic intention recognition. *Journal of Intelligent and Robotic Systems*, 45(1):31 – 52, 2006. [pp.]

Tegart. Smart technology for healthy ageing. Technical report, ATSE, 2010. [pp.]

The Repatriation Commission, Australia. Guide to the assessment of rates of veterans' pensions (garp), fifth edition. Technical report, Department of Veterans' Affairs, Canberra, 1998. [pp.]

S. Thrun, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Hähnel,

C. Rosenberg, N. Roy, J. Schulte, and D. Schulz. Minerva: A second-generation museum tour-guide robot. In *the IEEE International Conference on Robotics and Automation*, volume 3, pages 1999 – 2005, 1999. [pp.]

United Nations. United nations 2006 world populations prospects: The 2006 revision. Technical report, United Nations, New York, 2006. [pp.]

M. F. Valstar and M. Pantic. Combined support vector machines and hidden markov models for modeling facial action temporal dynamics. In *IEEE international Conference on Human-computer Interaction*, pages 118 – 127, 2007. [pp.]

V. N. Vapnik. *Statistical learning theory*. Wiley, 1998. [pp.]

Veterans Affairs Canada. Instructions and table of disabilities for the guidance of persons making assessments of the extent of disabilities. Technical report, Minister of Veterans Affairs Canada, 2006. [pp.]

Y. Wang, P. Sabzmeydani, and G. Mori. Semi-latent dirichlet allocation: A hierarchical model for human action recognition. In *Proceedings of the 2Nd Conference on Human Motion: Understanding, Modeling, Capture and Animation*, pages 240 – 254, 2007. [pp.]

G. Wasson, P. Sheth, C. Huang, A. Ledoux, and M. Alwan. A physics-based model for predicting user intent in shared-control pedestrian mobility aids. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 2, pages 1914 – 1919, 2004. [pp.]

G. Wasson, P. Sheth, C. Huang, and M. Alwan. Intelligent mobility aids for the elderly. In Majd Alwan and Robin A. Felder, editors, *Eldercare Technology for Clinical Practitioners*, pages 53 – 76. Humana Press, 2008. [pp.]

World Health Organisation. A glossary of terms for community health care services for older persons. Technical Report Volume 5, WHO Centre for Health Development, Ageing and Health, 2004. [pp.]

M. Zhang and A. A. Sawchuk. Motion primitive-based human activity recognition using a bag-of-features approach. In *ACM SIGHIT International Health Informatics Symposium (IHI)*, pages 631–640, 2012. [pp.]

C. Zhu, Q. Cheng, and W. Sheng. Human intention recognition in smart assisted living systems using a hierarchical hidden markov model. In *IEEE International Conference on Automation Science and Engineering*, pages 253 – 258, 2008. [pp.]

C. Zong, M. Chetouani, and A. Tapus. Automatic gait characterization for a mobility assistance system. In *11th International Conference on Control, Automation, Robotics and Vision*, pages 473 – 478, 2010. [pp.]