ASSISTIVE INTELLIGENT ENVIRONMENTS FOR AUTOMATIC HEALTH MONITORING

Daniel H. Wilson

CMU-RI-TR-05-42

Robotics Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

September 2005

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Thesis Committee:
Chris Atkeson, Chair
Andrew Moore
Jeff Schneider
Scott Hudson, HCII

Abstract

As people grow older, they depend more heavily upon outside support for health assessment and medical care. The current healthcare infrastructure in America is widely considered to be inadequate to meet the needs of an increasingly older population. One solution, called *aging in place*, is to ensure that the elderly can live safely and independently in their own homes for as long as possible. *Automatic health monitoring* is a technological approach which helps people age in place by continuously providing key information to caregivers.

In this thesis, we explore automatic health monitoring on several levels. First, we conduct a two-phased formative study to examine the work practices of professionals who currently perform in-home monitoring for elderly clients. With these findings in mind, we introduce the simultaneous tracking and activity recognition (STAR) problem, whose solution provides vital information for automatic in-home health monitoring. We describe and evaluate a particle filter approach that uses data from simple sensors commonly found in home security systems to provide room-level tracking and activity recognition. Next, we introduce the "context-aware recognition survey," a novel data collection method that helps users label anonymous episodes of activity for use as training examples in a supervised learner. Finally, we introduce the k-Edits Viterbi algorithm, which works within a Bayesian framework to automatically rate routine activities and detect irregular patterns of behavior.

This thesis contributes to the field of automatic health monitoring through a combination of intensive background study, efficient approaches for location and activity inference, a novel unsupervised data collection technique, and a practical activity rating application.

Acknowledgements

I owe the most thanks to my advisor, Chris Atkeson. I was lucky to be paired with such an amiable and ruthlessly intelligent person. I also appreciate the support of the National Science Foundation, specifically the NSF IGERT Graduate Fellowship (DGE-0333420) and NSF grants IIS-0312991 and IIS-0205219.

I thank Paul Carpenter, my best friend and study partner from kindergarten through twentieth grade.

Thanks to my collaborators at Microsoft Research, the Palo Alto Research Center, Northrop Grumman, and Intel Research Seattle, especially Andy Wilson, Allison Woodruff, Paul Aoki, Daniel Morris, Matthai Philipose, and Sunny Consolvo.

Many thanks to Virginia Arrington, Suzanne Lyons Muth, and Diane Stidle for keeping a sharp eye on the details.

Thanks to the many graduate students who gave good advice, especially Michael Montemerlo, David Tolliver, and Vandi Verma.

Special thanks to my good friends Abraham Flaxman, Andy Grieshop, Laura Gonzalez, and Anna Camille Long. Thanks also to my friends at CMU over the years: Sonya Allin, Brenna Argall, Curt Bererton, Allison Bruce, Deepayan Chakrabarti, Laura Dabbish, Scott Davidoff, Dave Ferguson, James Fogarty, Anna Goldenberg, Nidhi Kalra, Bradley Malin, Nick Patronick, Mamta Puri, Patrick Riley, Nicoleta Serban, Rande Shern, Ricardo Silva, Jake Sprouse, and Yuriy Nevmyvaka.

I thank my grandparents for their pride, for pushing me to finish, and for spreading the word via CB radio.

Finally, I thank my parents, Pam and Dennis Wilson; my brother, David Wilson; and my friend Tanner Rogers, for being a solid home base in sunny Tulsa, Oklahoma.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

People aged 65 and older are the fastest growing segment of the U.S. population – set to double in the next two decades [8]. As people age, they depend more heavily upon outside support for health assessment and medical care. The current healthcare infrastructure is widely considered to be inadequate to meet the needs of an increasingly older population. One solution is to enable *aging in place*, in which elders live independently and safely in their own homes for as long as possible, i.e., avoiding the transition to a care facility. This approach helps keep the elderly population happy and socially connected, while reducing the strain on healthcare infrastructure [82].

In this thesis, we contribute to the field of automatic health monitoring by conducting a study of professionals who currently perform in-home health monitoring, devising algorithms that infer location and activities of multiple occupants in a home, providing a technique for easily collecting training examples of activity, and by devising algorithms which can spot deviations in patterns of routine activity.

1.1 Overview

Automatic health monitoring technology uses sensors and machine learning algorithms to automatically collect information about patients for use by caregivers. In-home health assessment is largely concerned with monitoring clients' activities of daily living (ADLs), a set of activities used by physicians to benchmark the physical and cognitive abilities of patients [64]. Studies have shown that pervasive monitoring of the elderly and those with disabilities can improve the accuracy of pharmacologic interventions, track illness progression, and lower caregiver stress levels [39]. Additionally, [112] has shown that movement patterns alone are an important indicator of cognitive function, depression, and social involvement among people with Alzheimer's disease.

1.1.1 The Activities of Daily Living Study

In this thesis, we begin by describing a two-phased formative study of the professionals who currently provide in-home health monitoring to elderly clients and those with disabilities. In the first phase, we interviewed five participants for over one hour each. We used the results from phase #1 to design a questionnaire which was distributed during phase #2 to 91 participants across the United States. By understanding the current practices and identifying the major challenges faced by the professionals who care for elders, we hoped to determine how technological innovation could help elders age in place while respecting the most rewarding aspects of the case manager's job, the elder's need for autonomy, and the needs of the elder's family and friends who might also be impacted by the technology.

1.1.2 Simultaneous Tracking & Activity Recognition

In the next chapter, we quantify the perceived needs and constraints provided by the ADL study in the simultaneous tracking and activity recognition (STAR) problem. Solving the STAR problem provides the information we deem essential for automatic health monitoring in the home, including: identifying people, tracking people as they move, and knowing what activities people are engaged in. More challenging goals include recognizing when people deviate from regular patterns of behavior and providing advice on how activities could have been performed better.

We describe a particle filter approach to the STAR problem which exploits information gathered by many *simple* sensors. The benefits of particle filters are paramount when solving the STAR problem in a home environment with several occupants and several hundred sensors. Most importantly, particle filters offer a sample-based approximation of probability densities that are too difficult to solve in closed form, e.g., the data association problem that exists when tracking multiple occupants. Particle filters are more likely to recover from tracking errors because they can approximate a large range of probability distributions, unlike Kalman filters which are limited to Gaussian distributions. Computation can be minimized by using resampling, which focuses resources on only the most promising hypotheses. Finally, the number of samples can be dynamically adjusted according to available computational resources.

Our approach predominantly uses information from anonymous, binary sensors, particularly those employed by security systems, such as motion detectors and contact switches. We call a sensor anonymous and binary when it can not directly identify people and at any given time it supplies a value of one or zero. This severely limited amount of information comes with minimal privacy, monetary, and computational cost, and can be used on a large scale in homes and businesses. We show that such sensors can be used to solve the STAR problem, telling us which rooms are occupied, counting the occupants in a room, identifying the occupants, tracking occupant movements, and recognizing occupant activities.

1.1.3 The Context-Aware Recognition Survey

In the next chapter, we describe a novel data collection method called the "context-aware recognition survey." Accurate models of activity require labeled examples of activity for training, and the same activities inevitably vary between households and between individuals. Collecting labeled examples of activities is vital for training supervised learners to recognize individual human activities.

We devised an unsupervised data collection method which helps users to label anonymous activity episodes by displaying contextual information gathered by ubiquitous sensors in a game-like computer program. We found that users were able to consistently and correctly label these training examples, even when the activities were performed by other people. Our approach allows anyone to label the data at any time, without requiring additional hardware beyond the original sensor infrastructure and without causing any additional interruption to daily routine.

1.1.4 The k-Edits Viterbi Algorithm

Finally, we explore an important application area – routine activity rating and detection of irregular behavior. A main concern for health care professionals is to recognize instantly when clients' needs have changed. Thus, an important component of any automatic health monitoring system should be to recognize deviations in routine behavior and to isolate potential causes.

We introduce the k-Edits Viterbi algorithm, a polynomial time algorithm that works within an HMM framework to provide maximal likelihood modifications to a sensor trace generated in an instrumented environment. The k-Edits Viterbi algorithm may operate either as an activity rater, in which an activity is "graded" and suggestions for improvement are made, or as an activity monitor, in which deviations are detected and the reasons for the deviation are provided. Our approach is designed to be inexpensive and credible; the algorithm uses a straightforward HMM, does not require a description of possible problems, and provides advice that is constructive, relevant, and justified.

1.2 Thesis Contributions

This research thoroughly explores the field of in-home health monitoring before defining and providing a solution to the simultaneous tracking and activity recognition (STAR) problem. Our approach fills an important gap in existing research in ubiquitous computing by using a sensory modality that has been largely ignored in favor of vision and audition. We introduce a data collection technique that is designed to work hand-in-hand with existing techniques, such as the experience sampling method (ESM) [13], to provide labeled training examples which are crucial for supervised learning algorithms. We derive a new algorithm that can be used within a hidden Markov model framework to rate routine activities and pinpoint deviations in routine activity.

This work makes several contributions to the area of automatic health monitoring and to the field of ubiquitous computing in general:

- We identify key requirements for future technology by conducting an in-depth study of the care professionals who represent the current state-of-the-art.
- We introduce algorithms that perform simultaneous tracking and activity recognition of multiple occupants in a home, leveraging knowledge of location to improve activity recognition and vice versa.
- Our approach predominantly utilizes information from simple, cost-effective sensors common to home security systems, making it feasible to inexpensively instrument or retro-fit entire buildings.
- Cost-effective tracking and activity recognition creates the possibility of fundamental new interfaces or ways of interacting with structures and information processing resources.
- A new data collection technique is introduced which works alone or in concert with existing techniques to help users provide training data to supervised learners.
- We derive an algorithm which can rate how well a routine activity was performed and suggest improvements or detect possible causes of irregular behavior.

In the next section, we describe a possible future scenario that encompasses the technology that we wish to make possible through this thesis work.

1.3 Scenario

A man has an elderly mother living alone one hour away. Last week she knocked the phone off the hook and was unavailable for an entire day. The man walks into a hardware store and emerges with a large brown box. It contains several dozen nondescript, quarter-sized sensors that stick to any surface. Following directions, the man attaches the simple sensors to doors, drawers, and chairs in his mother's home. He pulls out a CD-ROM and installs software on a personal computer and plugs a device into a USB port. The software instructs him to perform a quick walk-through of the house, touching every sensor. Later that week the man logs onto the Internet, types a password, and is presented with ten multiple choice questions. In each question he is shown a series of images representing sensors that were touched and is asked to choose which activity was happening. One week later the man checks to see that his mother has been cooking and eating meals. One month later he checks whether her activity levels are steady. The system reports that activity levels are abnormally low today. He calls and finds that his mother seems to be coming down with the flu.

1.4 Thesis Layout

This thesis is organized as follows: In chapter 2 we present results from a study of the professionals who routinely perform in-home health monitoring. In chapter 3 we define the simultaneous tracking and activity recognition problem as a key component of automatic health monitoring and describe our particle filter-based solution. In chapter 4 we present a data collection approach for producing training data for our learner. In chapter 5 we describe a practical application that provides activity rating and/or detection of irregular behavior. Finally, in chapter 6 we conclude.

Chapter 2

In-Home Monitoring: A Study of Case Managers

In this chapter, we describe "The Activities of Daily Living Study," a two-phased formative study designed to examine the work practices of professionals who perform in-home monitoring of elders' activities of daily living (ADLs).[1] As a result of a swelling elderly population, there is a demand for technology that can assist caregivers and augment the elders' ability to age in place. Supporting elderly adults' preference to live independently at home, i.e., to age in place, can keep elders happy and forestall the transition to costly care-giving facilities [107]. By understanding the current practices and identifying the major challenges faced by professional caregivers, we hope to determine how technological innovation could help elders age in place while respecting the most rewarding aspects of the case manager's job, the elder's need for autonomy, and the needs of the elder's family and friends who might also be impacted by the technology.

[1] This chapter is a revised version of the technical report [118] and this work was performed in collaboration with researchers at Intel Research Seattle.

2.1 Introduction

Keeping elders safe and healthy in their own homes is accomplished with the help of caregivers – often the elder's family, friends, and neighbors. When professional care is necessary, it is commonly managed by individuals called case managers (CMs), professionals (not necessarily nurses) who interact with the elder, or client, at home and via phone calls, to assess mental and physical status and to arrange for necessary services to fill any gaps in ability (e.g., personal aides, assistive equipment, or cleaning services). During monitoring visits, a CM tracks the abilities of the client and is responsible for introducing new equipment (e.g., walkers, grab bars, etc.) and services (e.g., personal care assistants, home-delivered meals, etc.) as needed. For a CM, a significant part of assessment is captured by monitoring the client's performance of a specific set of activities – called Activities of Daily Living (ADLs). The Katz index of Activities of Daily Living was developed in 1963 as a measure of overall cognitive function and physical abilities; it includes activities such as bathing, preparing food, getting dressed, grooming, and eating meals [64].

The use of technology to automatically monitor ADLs has become a serious research focus worldwide [12, 21, 27, 28, 30, 45, 46, 67, 80, 92, 113]. However, at least in the fields of ubiquitous computing and human-computer interaction, there has been relatively little exploration of how information about ADLs in the home is currently collected and used by case managers to help elders age in place.

To begin our exploration, we conducted a two-phased formative study in which we examined the work practices of professionals in the U.S. who perform in-home monitoring of elders' activities of daily living. In phase #1, we conducted one-on-one semi-structured interviews with a small number of case managers (N=5) who maintain a combined caseload of 156 clients. In phase #2, we collected 91 paper-and-pencil questionnaires from case managers around the United States. The questionnaires covered many aspects of a case manager's job and were designed based on the results of phase #1.

In the next section, we discuss related work in the area of ADL monitoring and explain why the results of our study are both original and valuable. Next, we report on our methodology and results for both phases of this study. Finally, we describe the impact that these findings may have on several areas of research related to health assessment.

2.2 Related Work

Healthcare technology for the elderly has been a popular area of research, spawning the sub-discipline of "gerontechnology" [22]. Automatic monitoring of ADLs has been a common focus in gerontechnology; however, little research exists to motivate and guide such technology. Instead, research often relies on predictions of an exploding elderly population [4], studies that show the importance of ADLs for assessing functional ability [75], and/or "lessons learned" from actual hardware deployments [120, 54]. In this study, we examine the professionals who currently provide in-home health monitoring to determine which ADLs are most important to monitor, what obstacles stand in the way of collecting such information, and where technological innovation can (and cannot) meaningfully improve current practices.

Considerable effort has been devoted to applying technology to a wide variety of problems in the healthcare domain, including assistive robotic devices and remote health monitoring tools, i.e., tele-medicine or tele-health applications [94, 36, 101]. Researchers at the Georgia Institute of Technology use the "Aware Home" to study elders' reactions to such technology [74]. At MIT, researchers use the "PlaceLab" to develop design strategies for assistive technology in the home [57]. Other researchers have examined how older adults interact with technology by studying their cognitive abilities [73] and by examining how and where the elderly typically live [38]. For a recent survey of existing assistive technology see [47]. Previous research has been aimed at determining in a broad sense how and why to design technology for elders. In contrast, we restrict our scope to research directly related to monitoring elders who are living independently in their own homes.

Most research involving technology for monitoring ADLs is motivated at some point by information from government-sponsored, nationwide studies of the current and predicted characteristics of aging populations. Such sources include the United

States Census Bureau [111], the U.K. National Office of Statistics [110], the American Institute on Aging [4], the National Center for Health Statistics [114], and the Robert Wood Johnson Foundation [98]. These sources are valuable but non-specific, typically revealing information such as the size of elderly populations (in 1996 there were more than 44 million Americans over the age of 60), current and predicted life expectancy (in 2025 the projected number of Americans over 60 will be approximately 82 million, accounting for over 20% of the total U.S. population), and prevalence of disabilities and needs in these populations (20% of Americans over the age of 80 live in nursing homes). Information from these sources helps inspire technology for healthcare, but cannot adequately motivate specific technological design decisions and applications.

Existing literature strongly identifies the importance of activities of daily living (ADLs) such as bathing, toileting, and eating [64]; instrumental activities of daily living (IADLs) such as managing medication, maintaining a household, and preparing meals [68]; and/or enhanced activities of daily living (EADLs) such as using the Internet to connect to family and friends [99], as indicators of the physical and cognitive abilities of elderly individuals. The Intelligent Assistive Technology and Systems Lab at the University of Toronto found that ADL performance is important to know when designing technology for those suffering from Alzheimer disease [77]. At the Rose-Hulman Institute of Technology, researchers are developing simple tests to measure functional decline in elderly individuals, but identify ADLs and IADLs as the "gold standard" for measuring functional ability [75]. Policy makers at the American Association for Retired Persons (AARP) advocate the use of ADLs and IADLs for measuring levels of functional disability for establishing eligibility criteria for their programs and benefits [63]. Such research concretely identifies the usefulness of ADL measures for healthcare technologists. To our knowledge, no literature exists that explains in-depth how ADL information is (or should be) collected or which ADL information is most important to caregivers [81].

Instead, much research into ADL assessment technology is motivated retrospectively, in terms of "lessons learned" during actual prototype deployments. For example, researchers who are using simple sensors to detect the behavior of occupants in a typical home environment have provided lessons learned when instrumenting environments for the elderly [117, 12, 57, 92, 101]. For a partial survey of research concerning instrumented environments, see the Center for Aging Services and Technologies (CAST) technology demonstration program [28]. Other research covers topics such as convincing elders to complete ADLs [55] and closely monitoring ADLs to provide reminders (or cues) for how to properly complete ADL tasks [120, 34, 78]. These studies use ADLs to demonstrate technology, but without any guarantee that the chosen ADL is the best fit for the technology. For example, an automated cueing system might guide an elder through a hand-washing activity, however, it is not clear that hand washing is one of the ADLs that elders have difficulty performing. We believe that future research could benefit from knowing which ADLs are difficult or most important for a caregiver to collect; such knowledge could directly inform the design of automatic ADL monitoring applications.

2.3 Phase # 1: Interviews with Case Managers

In this section, we describe the methodology and results of phase #1 of the ADL study in which 5 case managers were interviewed in depth about their jobs. The interviews provided us with a rich description of what a CM's job is like and also helped us develop the survey instrument that was used in phase #2.

2.3.1 Study Methodology

Participants. Participants in phase #1 were 5 female case managers who were recruited by word of mouth at the case management agency where they were employed (located in Oklahoma, USA). Participants received a $20 gift certificate to a retail department store for participating. They ranged in age from 32 to 52 with years on the job varying from 2 to 10. Every participant had the job title "case manager." Two participants were also registered nurses (RNs), and a third was both the case management supervisor and an RN, and managed 13 case managers. Each participant routinely visits elderly clients to collect information about ADLs, with three CMs holding a case load of 50 clients each, the CM/RN holding a case load of 3 clients (with frequent skilled nursing visits to many others), and the CM supervisor holding a case load of 3 clients.

Sessions. Each participant engaged in a one-on-one session that lasted approximately 90 minutes. The session began with a short paper-and-pencil based questionnaire, followed by a semi-structured interview conducted by one of the study researchers. The questionnaire and interview are briefly described next.

Questionnaire. Participants spent approximately 15 minutes at the beginning of their sessions completing a paper-and-pencil based questionnaire about their age, job title, job experience, and on-the-job use of technology (e.g., cell phone, email, answering machine, etc.). See Appendix A for a blank copy of this questionnaire; see Appendix B for the aggregate results.

Interview. Semi-structured interviews were then conducted with each participant. All interviews lasted approximately 75 minutes each and were conducted in the home office of one of the researchers. Audio from the interviews was recorded on a digital recording device and later used to generate interview transcripts. Participants were asked questions in four main areas: 1) general job duties, 2) visiting a client in his/her home, 3) ADL forms and the flow of information, and 4) opinions about the use of automatic ADL monitoring technology. To preserve generality, we did not mention a specific technological approach. See Appendix C for the interview guide.

Data analysis. A code book was developed by the researcher who conducted the interviews to describe significant results and areas of possible interviewer bias. The researcher then used these codes to categorize the qualitative data contained in the interview transcripts. The code book was first validated by another of the study researchers and further validated by a third party researcher from outside the project, who independently coded the data using the code book. Code agreement between the study researchers and the independent coder exceeded 85%.

2.3.2 Results

During the interviews, shared themes emerged between every case manager, and we also identified a mix of promising opportunities and challenges for future technology.

In this section, we first describe the job responsibilities shared by all participants and then identify opportunities for technological innovation, before examining a series of challenges facing future technologies.

Job Description

Some common themes emerged from all interviewees. These themes describe job duties, technology use, and interactions with clients (including initial assessment, face-to-face home monitoring visits, and phone monitors).

Job duties. Case managers (CMs) work for the AGENCY, which receives money from the state government through a Medicaid program for every client who is on the PLAN.[2] Although most clients live at home, they may live with their families or in senior housing, but not in a nursing home. The AGENCY is reimbursed per client depending on the needs of that client, which may include skilled nursing, case management, or personal care services. The state government, along with the Health Care Authority, determine the reimbursement rates. These entities also govern the policies that the Long Term Care Authority (LTCA) sets forth. These policies and requirements are regularly audited at each AGENCY to determine that each client is being monitored and that their needs are properly being met.

A case manager does not require any medical training, however, they do require a degree in a social or nursing field – they are often trained as registered nurses (RNs) or licensed practical nurses (LPNs). The CM's job is to keep clients living safely and independently at home by frequent monitoring and assessing of the client's ability to complete ADLs. If the client is unable to safely complete ADLs and/or she needs assistance to properly complete the ADLs, then the CM will assess the need, complete a plan of care, submit for certification by LTCA, and monitor that the services are initiated. For example, a client may no longer be able to safely complete bathing, so the CM must initiate the proper paperwork so that a staffing coordinator may assign a personal care assistant (PCA) to aid the client with bathing. The CM determines the needs of the client, but it is the staffing coordinator, and not the CM, who actually works out the logistics of providing the care.

[2] Actual agency and plan names have been omitted to preserve confidentiality.

Formally speaking, the CM's goal is to keep clients safe at home with services from the state Medicaid PLAN, which is a state government-mandated program. Program policy mandates that CM's must make home visits and telephone calls to monitor ADLs and to reassess the client's needs as her health and circumstances change. Assessment includes a Mental Status Questionnaire (MSQ) [60], which determines levels of dementia based on whether the client is able to adequately answer questions. The program will also supply a client with medical equipment to meet a safety need or a health need (e.g., adult diapers, meal delivery service, financial counseling, mental health counseling, cooking, housework, or skilled medical care). The CM must also complete standardized paperwork so that every visit, every test, and every service ordered is documented and available in the event of a state audit. This paperwork may be hand-written or computer generated, depending on the individual CM's preference, i.e., there is no AGENCY or Medicaid program rule that dictates whether the paperwork is to be completed electronically or by hand.

Client eligibility. Any client may apply for assistance from the state at no cost. Requirements include financial distress, medical need, and some level of existing informal support (i.e., friends, family, or other non-paid caregivers). It is important to note that the goal of the AGENCY is to safely supplement existing support infrastructure – not to replace it. Each client must require nursing home level of care to qualify for this state program. It may be more reasonable for the client to go to a care facility if the cost of keeping the client safe at home exceeds the cost of a care facility. Case managers do not take part in evaluating clients for eligibility to the program (this is determined by the Department of Human Services); instead, CMs determine the client's needs once the client has been determined to be eligible for the PLAN.

Use of technology. Every participant interviewed regularly used a cell phone to coordinate and schedule visits with clients. CMs are not required to use a computer to fill out the forms, as long as they are written legibly. Many participants used a desktop computer; none used a laptop. Although cell phones and desktop computers were provided by the AGENCY, CMs often used their own equipment as well. At the time of the interview, laptops were about to be allotted to CMs by the AGENCY.

Initial Assessment. Upon acceptance to the program, a client must proceed through several steps. First, a registered nurse (RN) and a CM visit the home for an in-depth information exchange, called an "interdisciplinary team meeting" (IDT), with the client and any other outside support the client has. If the CM is also an RN, then only one person needs to make this visit. The CM will inspect the client's living environment to make sure that it is safe; she will also make arrangements to correct any unsafe situations (e.g., ordering grab bars for a slippery shower). Meanwhile, the RN will assess the client's health needs and determine what medical services may be necessary (e.g., diabetes care). After the visit, the CM assembles a PLAN of care that assimilates all of the information gathered during the initial assessment. The PLAN includes how often to visit, which services to order, and who is involved with the client (both formal and informal support) and how they assist with the client's needs. Each plan is reviewed by the CM Supervisor and sent to LTCA for final certification. CMs are responsible for initiating all services certified on the plan. Finally, the client is given a copy of her PLAN, and she must have it available for review for her CM during each home visit (or during an audit by the LTCA). The entire process, from notification of a new client, to initial assessment, to forming the PLAN and submitting it to the LTCA must be finished within 10 working days. The first initial contact is usually within 24 hours of receiving the assignment from LTCA, the home visit within 3-4 days, and certification must come within 10 days.

Regular Monitoring Visits. Each client is assigned to one CM, although CMs are allowed to visit each others' clients, e.g., to cover for a sick colleague. Policy dictates that a CM will contact the client monthly (via telephone or in-home visit), however, there are variables that determine if the client needs to be seen in person by the CM on a more frequent basis. The frequency of face-to-face visits versus telephone visits is determined by the risk level associated with each client. Risk level can be elevated for at-risk clients, e.g., if the MSQ score indicates moderate or severe dementia, if there is an unsafe environmental situation, if the health of the client is in jeopardy, or if a client is not adequately staffed although a need has been identified. High risk clients require a face-to-face visit every month, while lower risk clients only require a face-to-face visit every quarter (90 days). A home visit (i.e., face-to-face) usually consists of the CM driving to the client's home, sitting in a public area of the home (e.g., the living room or dining room), and reviewing the existing PLAN to determine if all of the client's needs are being met; the CM must also document the results of the in-home visit. Often, a CM will quietly inspect the environment (e.g., a CM may excuse herself to use the restroom, and then inspect it for safety and cleanliness). If a client uses a personal care assistant, CMs are occasionally required to complete a supervisory visit of the personal care assistant as mandated by the State. On the other hand, RNs visit on a different schedule than CMs to monitor health problems, but mostly to ensure that medication (prescribed by doctors) is being used properly (e.g., to ensure that a client is not missing doses). If the CM is also an RN, then she can perform both tasks in the same visit. At the end of one year, a reassessment visit is performed which follows the same procedure as the interdisciplinary team meeting (IDT) described above. This cycle continues until the elder no longer requires care, permanently moves to a care facility, passes away, or otherwise becomes ineligible for the program (e.g., finances may improve or a relative may move in to provide more support).

Paperwork. All official paperwork is designed by the LTCA and is also subject to audit by the LTCA, so CMs must be stringent about never losing information. Client confidentiality is also strictly maintained among all healthcare providers. Any services ordered, home visits, telephone communications, or other interactions must be documented and stored in a main file in the AGENCY office. CMs keep copies of their forms in a "travel file," so they can keep up with their clients while out of the office. This is vital, because CMs are often assigned up to 50 clients. Pieces of the travel file may be assembled by office staff and usually include a blend of distilled high-level information (not just copies of dense information). For example, a travel file may contain a client's name, address, date of last visit, and current list of services. The main file is accessible by anyone in the office with a need for it, including other CMs, as well as government officials. The clients' physicians may request the information, although this rarely happens. At the initial assessment, a "release of Information" form is signed by the client (or the client's power of attorney); this form lists every person that the AGENCY may speak to about the client. If a person who is not on the release requests information about a client, they are directed to ask the client about the matter – this includes family members and friends that the client may not want any information released to.

Opportunities for Technological Solutions

We now discuss the challenges faced by case managers and how they believed technology might help. We extracted several key results from the interview transcripts. Discussion is limited to the participants' current practices and their initial thoughts on the benefits and drawbacks of using automatic in-home ADL assessment technologies.

Most & Least Rewarding Aspects of Job. Any future technology that is designed to help case managers should keep in mind which parts of the job are most rewarding for a CM and which are their least favorite. When asked "What is your favorite part of the job?" every participant had the same answer – visiting the home and interacting with the client. When asked "What is the hardest part of your job?" three participants directly complained about the amount of paperwork they are required to complete. One participant claimed, *"Paperwork is ridiculous. It's stupid! S-T-U-P-I-D."*

Three participants mentioned that they felt technology could not reproduce their jobs, citing the elderly clients' need for social contact. One participant noted that technology is not necessary (*"...to put a tool in, instead of a human."*). Instead, participants mentioned using the technology to confirm second-hand information, to keep track of plans for multiple clients (i.e., time management), and to keep clients' families involved with client care. Based on the participants' responses, we believe that technology to help CMs complete paperwork might be very useful and welcomed.

Assessment of ADLs. One participant aptly describes her job as "being a detective." A CM must piece together information from clients, their family members, hired personal care assistants (PCAs), nurses, and many others to assess the true abilities and safety of the client. Some of the main challenges reported include: 1) not being able to observe the client perform activities, 2) not being able to trust the client's reported activities, and 3) not being able to adequately inspect the client's environment to find problems. These difficulties are discussed in more detail in the following paragraphs.

Participants expressed interest in technology to quickly detect changes in normal routine to help them ascertain when an unscheduled visit might be appropriate, (*"If you saw a big change, then that would tell you that you need to get out there."*). Additionally, they thought that technology might help them learn about things that may have been missed during visits (*"It would help identify the things that maybe we don't pick up on."*).

Auditing. A CM must decipher conflicting information from her experience in the home and conversations with clients, PCAs, and informal support. Elderly clients may exhibit reporting bias based on a desire to gain services or a fear of losing them (*"If I ask them if they've fallen, they may not tell me that they have, because they are afraid of going to a nursing home."*). The majority of monitoring visits are conducted by phone, where information is difficult to verify (*"All we go on mainly, especially in phone monitors, is what our consumers* [clients] *are telling us – which could be total B.S."*). In addition, PCAs work without direct supervision and CMs suspect that PCAs may exaggerate their hours (*"Aides* [i.e., PCAs] *lie because we will never find out."*). Finally, although informal support is usually described as being helpful, they are not always able to communicate freely. One participant says, *"Sometimes the family member doesn't want them* [the client] *to know that they are telling on them* [the client]."

Participants expressed interest in using technology to "sort out the facts." Two participants mentioned using technology to find out why the client isn't telling the truth e.g., maybe they have dementia. Three participants described using technology to verify that PCAs are really doing their jobs (*"The aide was sitting there watching All My Children, instead of working."*). Finally, while every participant mentioned that clients are not always truthful, one participant suggested that technology could be used to hold clients accountable for when they are untruthful, stating that then *"...they'd* [clients] *be less likely to not tell you the truth."*

Family involvement. Three participants mentioned that ADL information would be useful to the client's family members, to keep them involved in supporting the client. These non-paid caregivers, which the CMs call informal support, are often in tune with the day-to-day life of an elder, and their involvement and knowledge are very important to a CM (*"We like to visit with them* [informal support] *separately* [i.e., without the client present] *sometimes, okay, to get some adequate information."*). For example, informal support may notify CMs when there is a problem (e.g., the elder has been hospitalized). On the other hand, one participant noted that informal support is sometimes less informed, because some questions are too embarrassing for family members to ask (*"You aren't going to ask your mother if she took a shower."*).

Two participants pointed out privacy issues that monitoring technology could possibly allay: 1) minimizing the number of strangers who must enter the client's home, (*"...strangers are going to come into your home and all of the sudden ask you to remember things?"*) and 2) side-stepping certain awkward questions (*"...it is harder to ask the older gentlemen about incontinence."*). Monitoring technology could collect sensitive information in a more impersonal fashion, keeping informal and formal support better informed and more involved.

Scheduling. The three participants who work as full-time CMs reported an average caseload of 50 clients (Range=45-60). The CM who was also an RN reported a case load of 3 clients, but also made 20 skilled nursing visits per week to a variety of other clients. Every participant reported scheduling home visits based on hard deadlines imposed by state requirements (usually quarterly visits), and by the geographical distribution of cases (clients who live close together are easier to visit on the same day). The scheduling process was described as volatile, however, because elders' medical conditions often change suddenly (*"We're required to see them quarterly, but most often it's more than that because their needs change."*). Additionally, the day-to-day life of the client can affect scheduling (*"...I visited somebody three times because they are being evicted from their house..."*).

Four out of five participants expressed interest in using up-to-the-minute ADL information to improve scheduling (*"You might prioritize which ones* [clients] *you did home visits for."*). One participant suggested that technology could be used to set client risk levels (which dictate visit frequency) according to how compliant clients are with their PLAN goals. In addition, three participants described employing such technology to "fill in the blanks" between monitors, so that scheduling can be based on client need. One participant says, *"...if you see somebody that's a high priority...they've declined or started wandering or started falling more. You'd put them at the top of the list to go see."*

Monitoring functional decline. Several participants discussed how the abilities of elders decline over time (*"Usually in the beginning, a consumer* [client] *requires less assistance, but as they are on the program* [the PLAN] *three or four years, they are increasing."*). Participants described ADL monitoring as the main mechanism for tracking functional decline (*"Their ability to do things changes – mostly they deteriorate."*). Specifically, the CM is responsible for ordering additional services to replace lost functionality (e.g., ordering home-delivered meals for someone who is no longer able to cook).

One participant described using monitoring technology to *"watch them* [clients] *over time,"* to determine when additional services may be needed; the participant claimed, *"In the beginning their whole service plan is a low cost plan. Well, as they age...their plan becomes more expensive."* Another participant thought that technology could help the CM to *"see they've* [clients] *declined or they started wandering or started falling more."* The CM supervisor mentioned using the technology to monitor the high-level status of every client that is supervised by one of her 13 case managers (*"Well, it would be a lot easier than having to read through thirteen case managers' notes every month."*).

Challenges for Future Technology

Several interview questions were focused on case managers' envisioned use of advanced technology to assist with their daily job tasks. In this section, we explore the issues raised by case managers as they imagined using an unnamed, non-specific future technology capable of automatically assessing and disseminating information about clients' ADLs.

Concern over job loss. Although it was not a part of the interview guide, four participants expressed concern that future technology would put them out of a job (*"You have a computer do our job so we don't get paid?"*). However, one participant – the CM supervisor – expressed interest in using the technology to increase caseloads. Participants were enthusiastic about using technology for a variety of purposes, however, most were emphatic that technology not replace face-to-face contact between the CM and her client. To allay these concerns it is important that new technology respect the rewarding aspects of a CMs job, i.e., not try to replace the monitoring visits.

Acceptance issues. Most participants did not believe that their clients would accept an automatic ADL monitoring technology in their homes. Several participants made suggestions for making a system more palatable. They stressed that the system not be seen as a "tattle-tale," but that the technology should be seen as helping the client live at home independently as long as possible. One participant said, *"if it's* [the technology] *gonna be tracking every move I make, if it's gonna be telling them whether I'm going to have to go the nursing home or not...that could be a big issue."* Time and again, participants pointed out the negative association that elders have for the nursing home, and how strongly elders are motivated to avoid it (*"They think it's the end of the line; going to a nursing home means you are dead."*). Thus, successful acceptance of technology was directly linked to elders' understanding the technology's purpose of keeping them living at home.

Privacy issues. All five participants mentioned privacy issues as drawbacks to an in-home monitoring technology, and two participants described privacy issues as the main drawback of such a system. One participant mentioned privacy issues linked to cultural communities; *"Where is their privacy?"* says the participant, *"...the African-American community, they don't want you to know all their business."* Another participant was more optimistic, saying *"Some people are kind of fussy about their privacy, but most of the elderly want to tell you the daily things they are doing. They want to feel important."* Currently, information collected by CMs is accessible to other CMs, some office staff, physicians, the government officials who oversee the program, and to individuals who are specifically indicated by the elder. Several participants suggested that the current system of using "release of information" forms should also be used by the technology to determine access to data collected by a monitoring system. Obviously, respecting privacy is crucial for technology acceptance.

Delivering information. Participants had a range of familiarity with technology. Although every participant routinely used a cell phone, experience with computers varied (e.g., some CMs fill out forms by hand, while others use desktop computers). None of the participants carried a laptop or PDA to home visits, which perhaps explains why there was no mention of receiving information in emails or on PDAs. Of the four participants who answered this question, two were comfortable with receiving information on a computer or in a computer printout, another liked the idea of receiving information about clients on a web site, and one was most comfortable receiving the information via a phone call.

2.4 Phase # 2: Questionnaire with Case Managers

We used the results of the interviews to design a questionnaire that we distributed to case managers from four states around the U.S. We received 91 completed paper and pencil questionnaires. Questions focused on job duties, home environments, home visits, phone visits, reporting bias, visit scheduling, paperwork, and care networks. In this section, we describe the study methodology and results of phase #2.

2.4.1 Study Methodology

Participants. Participants in phase #2 were 91 adult individuals employed as case managers. Participants were recruited from four different case management agencies in the greater Seattle, Washington metropolitan area (N=49), the state of Oklahoma (N=22), the greater Pittsburgh, Pennsylvania area (N=15), and the state of Georgia (N=5). 12% of participants were male, and 88% were female. Participants were between the ages of 22 and 68 years (M=46, SD=10.49). Participants reported that they had worked at similar jobs for 0 to 35 years, with the mean number of years worked at similar jobs being 11.45 (SD = 8.5). Although exact job titles varied between agencies, all participants had a similar job description: to monitor clients' activities of daily living to help their clients live independently and safely. However, 15 of the participants worked with clients who lived in assisted living facilities and therefore did not technically make home visits.

Procedure. Each case management agency was approached separately, and participants were selected by word of mouth within their respective agencies, not directly by the study researchers; participation was voluntary. Questionnaires and incentives were distributed according to the following method: 1) a case management agency was selected, 2) a "point-person" for the agency was contacted, 3) questionnaires and incentives were delivered to the point-person, and 4) the point-person oversaw distribution and collection of the questionnaires among participants, administered the incentives, and returned the questionnaires to the study researchers. With one exception, questionnaires were distributed on a first-come first-served basis and collected by the point-person over the next few days. In one case, every questionnaire was distributed during a group meeting and questionnaires were filled out and returned simultaneously. Each participant received a $20 gift certificate to a coffee shop or department store (i.e., the incentive) in return for filling out a questionnaire. Participants filled out the questionnaires unsupervised and took from 30 minutes to two weeks to return the completed questionnaires to their point-person. As each questionnaire was collected, the point-person disbursed the incentive. Afterwards, the point-person returned the completed questionnaires to the researchers by mail or pickup.

Measures. The questionnaire was designed to take approximately 30 minutes to complete. Questionnaires were completed separately by each participant. Questionnaires were anonymous (containing no personally identifying information), however, each questionnaire came with a detachable information sheet for participants who were amenable to being contacted for future studies. The questionnaire focused on five main areas of case managers' jobs, based on results from the phase #1 interviews. Participants were asked questions in the following areas: 1) general job duties, 2) visiting a client in his/her home, 3) ADL forms and the flow of information, 4) clients' social networks, and 5) opinions about possible ADL-monitoring technologies. As with phase #1, questions did not suggest a specific technological approach. Many

Technology	Frequently	Occasionally	Rarely	Never	Don't have	What's that?
eMail	71%	18%	5%	2%	4%	0%
Internet	67%	23%	7%	3%	0%	0%
Voice mail	88%	9%	2%	1%	0%	0%
Electronic calendar	13%	11%	21%	36%	14%	4%
Instant messaging	10%	16%	26%	43%	3%	1%
Text messaging	15%	24%	24%	32%	2%	2%
OnStar	1%	2%	3%	41%	40%	13%
Computer / Laptop	83%	11%	1%	2%	2%	0%
Answering machine	80%	3%	3%	4%	9%	0%
Cell phone	86%	10%	1%	1%	2%	0%
PDA	9%	3%	10%	28%	47%	3%
MP3 Player	1%	3%	4%	31%	53%	8%

Table 2.1: Technology use among CMs from phase #2.

questions were multiple choice, but several were open-ended. Multiple choice questions were often accompanied by additional "fill in the blanks" for participants with insights that might not conform to the response options. To perform a quantitative analysis on the open-ended and "fill in the blanks" responses, a simple code book was developed and the write-in responses were coded as appropriate.

2.4.2 Results

Participants and technology use. The majority of participants reported using voicemail frequently (88%), a cell phone frequently (86%), a computer or laptop frequently (83%), email frequently (71%), and the internet frequently (67%). The least frequently used technologies were an MP3 player (84% never or don't have one), OnStar (81% never or don't have one), a PDA (75% never or don't have one), instant messaging (69% rarely or never), and an electronic calendar (57% rarely or never). See Table 2.1 for all results.

Specialized training. 39% of participants were RNs, 6% were LPNs, 23% were physical therapists, and 3% were occupational therapists. 43% received some sort of specialized geriatric training.

Monitoring clients. Participants reported regularly visiting a mean of 2.38 (SD=3.36) cities or counties for their job, and managing a mean of 35.38 (SD=31.36) clients at a time on average.[3] Participants reported making a mean of 2.33 (SD=2.20) home monitoring visits per day, and a mean of 1.42 (SD=2.11) monitoring phone calls per day. The questionnaire also asked participants to rank how much time they spent on each of four activities (home monitoring visits, paperwork, time in the office not doing paperwork, and traveling to and from home visits). Results, in order of where most of their time on the job is typically spent to least, were:

- home monitoring visits (49% of CMs ranked this most time consuming)
- paperwork (44% ranked this 2nd most time consuming)
- traveling to and from home visits (46% ranked this 3rd most time consuming)
- time in the office not on paperwork (51% ranked this least time consuming)

Main goal, challenge, and reward of job. When asked the open-ended question, "What is the main goal of your job?" the most common categories of responses were: keeping patients living in their homes or living independently (36%), patient (or client) safety (32%), patient care or health (24%), and developing plans or coordinating services (14%). When asked an open-ended question about what the most challenging aspect of their job was, the most common response categories were paperwork (21%), time management (14%), coordinating many services (14%), noncompliant patients (10%), and motivating clients or convincing them to accept services (10%). When asked what the most rewarding aspect of the job was, the most common response categories were improving patient quality of life or helping patients (40%), interacting with clients (21%), and helping patients be independent (18%).

Arranging services. The majority of participants reported that they frequently or occasionally arranged for the following services or items for their clients:

- visits from nurses (59% frequently, 30% occasionally)

[3] These results include responses from "clinical managers" and similarly titled case managers, whose responses may include the client loads of an entire group of case managers, not just their own.

- PCAs to help with bathing (56% frequently)
- shower chairs or grab bars (49% frequently, 31% occasionally)
- assistive walking devices (44% frequently, 42% occasionally)
- PCAs to help with cleaning (36% frequently, 32% occasionally)
- PCAs to help with cooking (31% frequently, 34% occasionally)
- adult diapers (31% frequently, 28% occasionally)
- home-delivered meals (30% frequently, 31% occasionally)
- mental health counseling (17% frequently, 41% occasionally)
- transportation (16% frequently, 49% occasionally)
- hospice visits (13% frequently, 44% occasionally)

It was more rare for participants to arrange for financial counseling for their clients (40% reported rarely doing this), and while 40% reported occasionally calling adult protective services for their clients, 53.9% reported doing this rarely or never (with 6.1% reporting frequently).[4]

Clients. The age range of clients seen by participants was 39-95 years, with the average age of clients as 69 years of age. The majority of participants (78%) reported that all or most of their clients were above the age of 65 and the other 22% reported that some of their clients were above the age of 65. CMs reported that their clients under the age of 65 were most likely to have multiple sclerosis (21%), cancer (20%), heart problems (17%), diabetes (14%), joint replacement (14%), or quadriplegia or paraplegia (12%).

The majority of participants reported that frailty (79%) and diabetes (56%) were very common problems among their clients. Other problems were reported by the majority of participants to be somewhat or very common, including:

[4]Adult protective services (APS) may help clients by prosecuting criminally negligent care providers or moving clients to safe surroundings.

- food allergies (65%)
- deafness (63%)
- blindness (56%)
- nutritional deficiency (54%)
- moderate dementia (53%)
- obesity (53%)
- mild dementia (50%)
- incontinence (49%)

Muteness (79% not very) and severe dementia (42% somewhat, 38% not very) were reported to be somewhat or not very common by the majority of participants. The majority (58%) reported that few or none of their clients were bedridden, while 39% indicated that some were, and 3% that many or most were. The majority of participants (78%) also reported that some or most of their clients use a wheelchair, while 91% reported that some or most of their clients use a cane, and 43% reported that some clients use a motorized scooter. Participants indicated that it is rare for clients to share assistive equipment with others (e.g., clients do not typically share a scooter with someone else in their household). 57% of participants indicated that their clients frequently wear shoes during a home monitoring visit.

Participants reported that most of their clients live in their own houses (57%) and apartments (49%) and that some live in family houses (60%), i.e., the home of a family member. Few to none of their clients lived in assisted living facilities (reported by 66%), shared houses (75%), or shared apartments (83%). 42% of participants reported that most of their clients lived alone, and 48% reported that some of their clients lived alone. 68% indicated that some of their clients had cats or indoor dogs. When clients did not live alone, participants wrote in that the client typically lived with only 1 other person (48%), 1 or 2 other people (21%), 2 other people (19%), or more than 2 people (13%).[5]

Health Information. The majority of participants reported that they frequently collected information about their clients from other professionals, including nurses (58%), personal care assistants (52%), doctors (39%), and other CMs (37%), as well as from the client's family members (61%). The majority reported that they rarely or never got information about clients from neighbors (68%) or landlords (81%). Overall, participants reported that they frequently felt like the information they got from these other parties was accurate, with information from nurses (90%), doctors (84%), and other CMs (78%) being ranked most accurate, followed by information from family members (76%) and PCA's (76%). Information from neighbors (48%) and landlords (60%) was most often ranked as being only occasionally accurate among participants who indicated that they interact with neighbors or landlords about the client.

When asked how they conduct a home monitoring visit with a client who they have difficulty communicating with, participants reported that they observe the client (47%), ask the primary caregiver (35%), and observe the environment (31%) to glean information about the client's ADLs. The majority of participants reported occasionally (45%) or rarely (37%) using an interpreter with a client. 25% reported using demonstrations or gestures when they have difficulty communicating with a client to help determine how well the client is performing ADLs.

When asked who contributes most to the typical client's care, most participants reported that a family member or child was most involved in providing care (62%), followed by a hired caregiver (32%) or the client's spouse (28%).[6] The majority of participants (81%) reported that they frequently interact with the individual most involved in the care of the client (15% occasionally). Participants reported interacting with personal care assistants (81% face-to-face, 68% telephone), family members (78% face-to-face, 82% telephone), and other case managers (51% face-to-face, 77% telephone) through telephone calls and face-to-face interactions, while they primarily

[5]This was an open-ended question, and participants did not always respond with whole numbers, e.g., "one or two other people."

[6]The percentages sum to over 100%, as 30% of CMs reported more than one individual contributes most to the client's care.

interact with nurses (47% face-to-face, 88% telephone) and doctors (14% face-to-face, 93% telephone) through telephone calls.

When asked who they believe their clients trust most with information about the clients' health, participants reported that clients trust doctors (34%), family members (25%), and nurses most (24%), followed by CMs (16%).

The Home Monitoring Visit. When visiting the client in his/her home, 72% of participants reported that they usually conduct the interview in the living room, while 30% reported using the kitchen, and 29% reported using the bedroom. The majority of participants also reported that they usually inspect the bathroom (66%), kitchen (55%), and bedroom (50%) during a home visit. 39% of participants reported that they are allowed to visit or inspect all rooms in the house during a home visit, while another 23% reported that they are not allowed to visit bedrooms belonging to people besides their client. Only 14% of participants reported that their job would be easier if they could visit rooms that they are not allowed to visit (including but not limited to the rooms of other household members).

When asked how often clients cancel or miss a visit, 54% indicated occasionally and 45% indicated that this rarely happens. 9% reported making unscheduled visits frequently, while the majority of participants indicated that they made unscheduled visits occasionally (38%) or rarely (33%). When asked why they make unscheduled visits, the most common responses were: because they could not contact the patient on the phone (13%), because they were in the area or just to make friendly contact with the client (12%), and to make sure that the PCAs show up for work (10%). When scheduling a home visit, the overwhelming majority of participants reported that they always (41%) or frequently (29%) consider agency or state deadlines, as well as changes in a client's medical status (49% responded always, 42% responded frequently). Receiving phone calls from a client or a client's friend, family, doctor, or nurse was reported to occasionally or frequently be a factor in scheduling, and being near a client's house was also reported to occasionally be considered when scheduling a visit. When asked if they had called 911 in the past 12 months because a client did not answer the door when the participant arrived for a scheduled visit, 80% reported that they had not, and 20% reported calling 911 one or more times, with a range of

Technology	Definitely	Maybe	Never	Unsure	What's that?	Already have
Emergency call buttons	77%	22%	0%	0%	0%	1%
Home security system	57%	39%	1%	2%	0%	1%
Cell phones	51%	42%	2%	1%	0%	3%
Internet	18%	66%	4%	8%	1%	3%
Computers / Laptops	16%	64%	6%	10%	1%	3%
Motion detectors	13%	58%	6%	22%	0%	1%
Microphones	8%	43%	18%	31%	0%	0%
Cameras	6%	51%	11%	31%	0%	1%
PDAs	5%	51%	12%	30%	2%	0%

Table 2.2: Participant feelings on clients' acceptance of technology.

0 – 10 (M=.43, SD=1.28).

Technology for Monitoring ADLs. When asked "If a magic genie could tell you every single activity that your client performed during the course of a day, would you still want to visit the house? Why or why not?" 97% (88 of 91) responded that they would still want to visit the house. The most common reasons given were to: provide face-to-face social contact (23%), directly observe the patient performing activities (20%), and see for myself/verify genie accuracy (13%).

When asked whether clients would allow various technologies in their homes, participants were mostly positive about their clients allowing emergency call buttons (77% definitely, 22% maybe), a home security system (57% definitely, 39% maybe), and cell phones (51% definitely, 42% maybe). Participants were slightly less positive about their clients allowing the Internet (18% definitely, 66% maybe), computers or laptops (16% definitely, 64% maybe), and motion detectors (13% definitely, 58% maybe). Participants were more wary of their clients allowing cameras (51% maybe, 11% never, 31% not sure), microphones (43% maybe, 17% never, 31% not sure), and Personal Digital Assistants (PDAs) (51% maybe, 12% never, 30% not sure). See Table 2.2 for all results.

When asked how participants would prefer to receive updates or reports on their client's condition (e.g., the type of information a system might provide about a client), phone calls were the most preferred (79%), followed by fax (42%), printout (41%), website (35%), video, (22%), and text message (21%).

When asked which ADL would be most valuable to have information about between home visits, the most common responses were:

- taking medications (23%)
- eating and drinking (17%)
- ambulation i.e., walking (14%)
- nutrition (14%)

When asked how important it is to know about various ADLs, participants reported the following as being very important:

- falling (99%)
- taking medication (93%)
- eating/nutrition (84%)
- toileting (82%)
- bathing (80%)
- getting out of bed (78%)
- grooming/hygiene (74%)
- dressing (68%)

Getting out of the house (56% very, 40% somewhat) and socializing (43% very, 52% somewhat) were both considered to be somewhat important, as were doing laundry, cooking, and cleaning. Yard work was the least important (36% somewhat, 36% not very, 13% not at all). Overall, participants reported that it was somewhat, not very, or not at all difficult to get information about these ADLs. However, a few of the activities were more frequently rated as very difficult to get information about, including: taking medication (11% very, 47% somewhat), falling (10% very, 46% somewhat), eating/nutrition (10% very, 53% somewhat), and socializing (9% very, 34% somewhat).

When asked what makes an ADL difficult to monitor, the most common responses were: the client was not honest (26%), the participant cannot observe the client completing the task (22%), and that the client is demented or agitated (10%). Overall, participants reported that clients are very or somewhat accurate in reporting on ADLs. For most ADLs, 20% or more of participants reported that clients were very accurate. However, the ADLs with the lowest reported accuracy were: grooming/hygiene (14% very accurate, 72% somewhat accurate, 11% not very accurate), taking medication (16% very, 72% somewhat, 10% not very), and eating/nutrition (16% very, 64% somewhat, 17% not very). Taking medication, falling, and eating/nutrition were the most often endorsed as being ADLs that the participants must frequently depend on information from people besides the client to assess. Participants also reported that they inspect the client's environment to get information about certain ADLs. In particular, most participants reported inspecting the environment to assess bathing (59% always), falling (58% always), grooming/hygiene (57% always), taking medication (57% always), and dressing (56% always).

When asked "If a magic genie could remind all of your clients to do one thing every day, what should the genie remind them of?" the most common responses were to: use medications correctly (57%), exercise (16%), and eat well (10%). When asked what one client behavior they would most like to change, the most common responses were: get more exercise (13%), follow the schedule or care plan (13%), ask for help and accept it (10%), and have more healthy habits (e.g., quit smoking) (10%).

Paperwork/Forms. 46% of participants have to fill out one form after a typical home monitoring visit, while 26% reported filling out two forms, 12% fill out three forms, and 16% fill out 4 or more forms. Participants reported referring to forms that they previously filled out a mean of two times per day (M=1.95), and referring to forms filled out by other CMs about one time per day (M=1.27). Participants indicated that home monitoring forms are most accessed by office staff (reported by

87% of the sample), other CMs (60%), government officials (54%), the client's nurse (53%), and insurance companies (41%). Fewer participants indicated that the forms were accessed by doctors (32%), PCAs (22%), the client's spouse (13%), or children (2%).

Participants reported spending between 0 and 210 minutes completing forms after a home visit (M=30.50, SD=31.29). 65% reported completing the forms in the client's home, while 42% completed forms in the office, and 24% reported completing forms in the car. 67% reported taking notes using paper and pencil, while 29% take notes using a laptop just after a home visit. 67% of case managers reported that they use a computer to fill out forms for home visits, and 33% reported not using a computer.[7]

2.5 Discussion

The results of this study may help researchers who are interested in developing technology for the automatic health assessment of elders, specifically automatic monitoring of ADLs. In this section, we discuss some derivative results from phase #2.

2.5.1 Understanding Case Managers' Motivations

Although all participants were case managers, training varied and included registered nurses (RNs), licensed practical nurses (LPNs), physical therapists, and specialized geriatric training. We found correlations between job background and a CM's main goal at work, most challenging aspect of work, and most rewarding aspect of work. We discuss these results next.

Job background and main goals. Results show that RN's are significantly less likely to say that the main goal of their job is "patient care" in general (r=-.40, p=.00), and are significantly more likely to report that "keeping the patient healthy" (r=.21, p=.05) is the primary goal of their job. Being an LPN was uncommon in the sample (only 5.7% of participants), but these individuals were significantly more likely

[7]For some case managers, the choice of whether or not to use a computer or laptop may be mandated by the agency for which they work.

to report that keeping patients living independently was the main goal of their job (r=.26, p=.01). On the other hand, physical therapists were less likely to report that keeping patients healthy was the main goal of their job (r=-.47, p=.00). Participants with specialized geriatric training were more likely to report that patient care was the main goal of their job (r=.24, p=.02), and were less likely to report that keeping patients living in their homes was their main goal (r=-.29, p=.01).

Job background and most challenging aspect of work. Being an RN, an LPN, or a physical therapist was not related to what the participants reported as being challenging. However, occupational therapists (only 3.4% of the sample) were significantly more likely to report that time management was the most challenging aspect of their job (r=.26, p=.01). Additionally, participants with specialized geriatric training were less likely to report that motivating clients or convincing them to accept services was the most challenging aspect of their job (r=-.22, p=.04).

Job background and most rewarding aspect of job. Being an RN is related to what participants reported as being the most rewarding aspect of their jobs. RNs are significantly more likely to report that improving the patient's quality of life is most rewarding (r=.36, p=.001), and are more likely to report that meeting client goals is most rewarding (r=.21, p=.05). Physical therapists were less likely to report that improving quality of life was the most rewarding aspect of their job (r=-.22, p=.04), and were more likely to report that helping a client be independent was the most rewarding aspect (r=.22, p=.04). Participants with specialized geriatric training were significantly less likely to report that meeting patient goals was the most rewarding aspect of their job (r=-.29, p=.01). All participants said that they would still visit clients, even if a technology could report every ADL, regardless of their background. However, we found a trend level association between reporting that the most rewarding aspect of the job is interacting with the client and reporting a desire to still visit clients in order to have face-to-face social interaction (r=.18, p=.10).

2.5.2 Interactions with Clients' Care Providers

We found a correlation between job background and how/who participants interact with during the course of their work. Depending on job background and years of experience, participants showed differences between who they collect information from and how much they trust that information.

Who participants interact with. RNs are significantly more likely to frequently collect information about clients from personal care assistants (r=.26, p=.01), family members (r=.32, p=.00), doctors (r=.28, p=.01), and other CMs (r=.34, p=.00). LPNs were less likely to collect information from family members (r=-.23, p=.03) and other CMs (r=-.22, p=.04). Physical therapists were less likely to collect information from family members (r=-.24, p=.03), and from landlords (r=-.24, p=.03). Having specialized geriatric training was related to less collection of information from doctors (r=-.23, p=.04).

Experience and PCAs. Results show that years of experience was positively correlated with a higher frequency of collecting information from personal care assistants (r=.23, p=.03). We believe that this may be due to the CMs finding the type of information that PCAs can provide more valuable as the CMs gain experience on the job, or due to years of experience correlating with more supervisory tasks related to PCAs.

Experience and trust. Overall, there were few significant associations between job and/or experience and how often participants felt that the information they get from others is accurate. Being an RN, LPN, or occupational therapist was not related to perception of accuracy of any sources of information, nor was years of experience. Physical therapists were less likely to think that information from landlords and neighbors was accurate (r=-.27, p=.01; r=-.23, p=.04). Having specialized geriatric training was related to a higher perception that information from neighbors was accurate (r=.32, p=.00).

2.5.3 Acceptance of Technology

We explored how a participant's own technology use predicted her estimation of her clients' acceptance of technology. To help answer this question, we created a technology acceptance score that sums across all the possible technologies that participants were asked to predict acceptance of. We also examined the technologies separately as individual variables.

In short, we found that a CM's own technology use did not predict how they felt their clients would accept technology. General technology use (question 4 in the questionnaire) and using a computer to complete forms for home visits (question 69) are not significantly correlated with predicted technology acceptance. Neither is age, years of experience, case load, nor degree title (e.g., RN, LPN, etc.). However, the participant's gender was a significant predictor. Women are more likely to report that their clients would allow technologies in their homes, including cameras and motion detectors (t(88)=1.80, p=.07). Although this finding is at the trend level, in a more complex model predicting technology optimism, gender emerged as the only significant predictor.

When participants were asked how they take notes during or just after a home monitoring visit (question 68), the 29% of the sample who reported using a laptop to take notes was more optimistic about technology acceptance by their clients, though this was a trend-level finding ($t(88)=1.66$, $p=.10$). We believe that these participants may have seen their clients interact with or react to technology by using a laptop in their clients' presence.

The city where the participant resides is also not related to their prediction of clients' technology acceptance. Despite this, we found that the city of residence is related to technology use, with Seattle participants reporting significantly higher levels of technology use than the other 3 locations ($t(88)=-2.45$, $p<.05$). We note that the use of laptops among participants in Seattle is not necessarily a personal choice, and may have been mandated by the agency that employs them.

We found that characteristics of clients also do not lead participants to be more or less optimistic about technologies being accepted. For example, average age of the clients, the number of clients above the age of 65, and the number of clients who live alone were not significantly related to prediction of technology acceptance.

It appears that optimism about whether clients would accept technology is a fairly independent construct. Though there are differences in technology use among the participants based on city of residence, neither city of residence nor a participant's own technology use is related to what they predict their clients would accept in a home-based monitoring system. Characteristics of the clients are also not related to predicted acceptance. It may be that the clients are perceived as an older generation who are less accepting of technology in general, regardless of where they live.

2.5.4 Time Spent Filling out Paperwork

We were interested in finding what might predict how much time is spent filling out forms. We discovered that a participant's age is positively correlated with the amount of time she spends filling out forms for a typical home monitoring visit. Older participants reported spending more time on paperwork (r=.28, p<.01). The average age of clients was not related to time spent completing paperwork, and neither were years of experience, degree title (e.g., RN), nor having specialized geriatric training. Using a laptop or computer to take notes or fill out forms versus using paper and pencil was also not predictive of perceived amount of time spent on paperwork.

2.6 Implications to Current Technology Research

The results of this study are directly applicable to several key areas of ubiquitous computing research. In this section, we will apply the information reported by interviewees from phase #1 and the questionnaire respondents from phase #2 with respect to people tracking, activity recognition, and the elder's care networks.

2.6.1 People Tracking

People tracking is a fundamental problem in ubiquitous computing and has been approached by many researchers, including the authors [117]. People tracking research is concerned with using information from sensors to estimate the location of some number of people at some granularity. Previous research with an elderly population has shown that location in the home can be an important indicator of social involvement, cognitive functioning, and physical activity and ability [39]. Our findings help describe the home environment (e.g., use/sharing of assistive equipment) and occupants (e.g., pets, number of occupants), information which can help researchers choose appropriate sensors and set appropriate tracking goals.

Location in the home. We found that most elderly occupants of a home will be mobile, as only 3% of questionnaire respondents reported that all or most of their clients were bedridden. All respondents identified that knowing whether a client gets out of bed in the morning was either very important (78%) or somewhat important (22%). 14% of respondents chose ambulation as the single most important ADL to know about. In addition, three interviewees mentioned that it is important to know whether a client moves around the house, so that they avoid "skin breakdown,"[8] which can be caused by remaining in the same place for too long (*"They don't move like we do. So they're sitting in one place all day long. They get skin breakdown."*). We hypothesize that room-level tracking could provide knowledge of where clients move in the home, which could reveal whether occupants are "up and about." Smaller than room-level granularity, or perhaps sensors to detect that the client has not gotten up from a single piece of furniture, may be required to determine that clients are not sitting in the same place for too long.

Scheduling. The majority of interviewees reported that simply knowing whether or not a client was at home could improve scheduling. Interviewees reported being very annoyed when clients are absent for a home monitoring visit, saying *"if nobody's home, it can be very irritating," "it makes me mad,"* and *"oh, you better be home if I've called you."* The majority of questionnaire respondents (54%) reported that their clients occasionally miss visits. Additionally, interviewees reported that they are required to make an emergency call to 911 if a client cannot be located for a scheduled home visit (*"Sometimes they've passed away. Sometimes they're asleep. Sometimes they have had a medical problem during the night and they're not able to get to the door."*). Questionnaire respondents reported making such emergency calls

[8] "Skin breakdown" is sometimes referred to as "bed sores" or "pressure sores"

only rarely, usually less than once per year (M=.43, SD=1.28), however, when asked how often calls to 911 are made, one interviewee states that *"one time is too often."* These findings indicate that home-level people tracking could improve scheduling by reducing the number of missed visits, thereby further reducing emergency calls to 911 as well as decreasing the worry and frustration associated with unaccounted for clients.

Hospital visits and socialization. Interviewees indicated that socialization is also an important indicator of overall functioning (*"it's good mental health not to sit up in a dark room all day long."*). 96% of respondents report that getting out of the house is very or somewhat important to know, and 95% reported that socializing was very or somewhat important to know about. In addition, two interviewees mentioned that they intentionally try to connect clients socially, *"We don't want them* [clients] *to be socially isolated. We want them to get out and go to church...get involved in functions...get outside their home."* Besides socializing, visiting the hospital is an outside-the-home activity that is important for a CM to know about. Three interviewees mentioned that the CM is often the last to know when a client visits the hospital (*"the hospital doesn't always let us know* [that a client has been hospitalized]."). We hypothesize that information gathered by a city-wide people-tracking system could be an important indicator of medical status and socialization.

Number of occupants at home. 42% of questionnaire respondents indicated that most of their clients live alone. Meanwhile, 88% of respondents indicated that when their clients live with others, they usually live with 2 or fewer people. This indicates that it is reasonable to expect 3 or fewer permanent occupants on average, not counting temporary visitors. There may often be extra people around; for example, most respondents (56%) mention that they frequently arrange for aides to visit their clients. People tracking technology should accommodate anywhere from one to half a dozen occupants at the same time.

Pets. Indoor pets are common, with 81% of respondents reporting that some or most clients had cats or indoor dogs. This indicates that researchers should expect to deal with constraints introduced by pets.

Instrumentation opportunities. Although shoes represent a key target for placement of wearable sensors, participants reported that clients only wear shoes at home about half the time, indicating that instrumented footwear will probably not be broadly successful for tracking in the home. Respondents indicated that assistive devices are often used by their clients; 45% report that all of their clients use canes (46% most), 24% report that most clients use a wheelchair (54% some), and 43% report that some clients use motorized scooters (5% most). Additionally, participants reported that clients almost never share assistive equipment with others. This indicates that identity may be associated with instrumented assistive devices. Heavily used assistive devices are promising locations for instrumentation, e.g., wheelchairs could sense weight or canes might detect obstacles or even general ambulation. On the other hand, these devices could introduce noise for other sensors, e.g., canes and wheelchairs cause occlusion for cameras, and metallic wheelchairs may interfere with RFID sensors. Of course, all of these opportunities assume that the CM has an accurate understanding of how the client uses her equipment.

Privacy constraints for sensor choice. Privacy issues, both actual and perceived, are crucial for researchers developing people tracking and other monitoring technologies to consider. Every interviewee mentioned privacy issues as drawbacks to an in-home monitoring technology (*"...but where is their* [the client's] *privacy?"*). When asked what technology they believed clients would accept, the majority of CM respondents replied that cameras, microphones, and PDAs would never be accepted or would only maybe be accepted. 22% respondents felt that motion detectors would definitely be accepted (78% maybe), while 57% felt that home security systems would definitely gain acceptance (39% maybe). Motion detectors are a common component of any home security system. This indicates that perceived privacy issues can be allayed by folding sensing technology into a larger, easier-to-understand package – such as a home security system. In addition, privacy issues may be mitigated if clients are aware of clear, immediate benefits accompanying new technology.

Very Important to Know		Very Difficult to Collect		Top Ten
Falling	99%	Taking medication	11%	Taking medication
Taking medication	93%	Eating/nutrition	10%	Falling
Eating/nutrition	84%	Falling	10%	Eating/nutrition
Toileting	82%	Socializing	9%	Toileting
Bathing	80%	Getting out of bed	7%	Getting out of bed
Getting out of bed	78%	Cooking	4%	Bathing
Grooming/hygiene	74%	Toileting	4%	Cooking
Dressing	68%	Bathing	3%	Socializing
Cooking	61%	Grooming/hygiene	3%	Grooming/hygiene
Leaving the house	56%	Dressing	3%	Dressing

Table 2.3: ADLs ranked by importance, difficulty to monitor, and top 10 most useful.

2.6.2 Activity Recognition

In-home performance of ADLs is a key measure by which CMs ensure that elderly clients are living safely and independently. One interviewee claimed, *"Completing ADLs and not completing ADLs tells a lot about somebody and how they are doing."* Many research projects have focused on automatic detection of ADLs [92, 77]. Findings from both phases of this study help describe several promising areas for future technology that automatically monitor ADLs.

Choosing which ADLs to monitor. We found several promising areas where technology could provide important information that is normally difficult to come by for a CM. Several ADLs were rated as being very important to collect information about, yet also difficult to get accurate information about. ADLs that were very important yet difficult to collect information about include: taking medication (also called "medication compliance"),[9] eating/nutrition, ambulation/activity level, falling, and socializing. When asked "If a magic genie could tell you everything about one of your client's activities' of daily living, which would be the most valuable?," the most common responses were taking medications (23%), eating/drinking behavior (17%), nutrition information (14%) and ambulation information (14%). See Table 2.3 for a

[9] One medical study estimates that 40% to 75% of elderly adults are non-compliant when taking medication [102].

breakdown of most important and difficult to collect ADLs, with a ranking of the top 10 ADLs from both categories.

Importance of grooming and hygiene. 83% of respondents indicated that client reports of grooming/hygiene were only somewhat or not very accurate. 89% of respondents reported always or frequently inspecting the environment (including the client's appearance) to determine information about grooming/hygiene. Every interviewee also reported inspecting the environment for this information. Clearly, CMs seem to rely on their own senses to gauge grooming/hygiene behavior, and 58% of respondents indicate that it is not very or not at all difficult to determine this information. Although grooming/hygiene information is easy to determine, our findings indicate that it may have an impact on "quality of job" for a CM. Two interviewees described messy homes or unclean clients as a drawback to home visits. One participant says that one of the hardest parts of her job is *"...just sometimes being disgusted by the way persons let themselves go,"* the same CM mentions, *"I've had consumers* [clients] *...I can't go to their house until I know they've taken a bath."* ADLs that are related to a clean client or a clean environment are not the most difficult to collect information about (uncleanliness is usually very apparent), but may have a positive impact the happiness of CMs and the quality of their home visits with clients.

Recognizing non-client activity. Elderly clients are only one of several classes of individuals who case managers must interact with during the course of their work. In many cases, interviewees mentioned that it would be useful to know the activities of informal support (e.g., spouses, relatives, visitors) and formal support, especially PCAs.[10] One interviewee reported that one of her main responsibilities is *"to see if the aide's doing what they're supposed to."* The same interviewee reported that client reports of aide performance are not always accurate, because *"some of our little consumers* [clients] *don't want to get the aides in trouble. So, they'll say the aides are doing* [work]." Another interviewee mentioned keeping track of informal support e.g., family members, as well, to make sure they are doing what they say they are

[10] We suspect the interviewees were referring to the activities related to the care of the client, and not all the day-to-day activities performed by informal and formal support.

doing. This interviewee went on to describe this as a difficult part of the job, *"You're basically calling them out on taking care of their family."* We hypothesize that it may be useful for activity recognition technology to not just focus on the elderly occupant of a home, but to keep better track of all aspects of her care, including monitoring the relevant activities of occupants and visitors (e.g., PCAs and family members).

Detecting cooperative activities. Interviewees reported that they often arranged for clients to receive help with common ADLs such as cooking meals, bathing, or dressing (*"...we* [arrange] *for the aide to cook with their assistance, but cook their* [the client's] *food the way they* [the clients] *cook stew or the way they* [the clients] *do a meatloaf."*). Questionnaire respondents also indicated that they frequently arrange for aides to help with bathing (56%), cleaning (36%), and cooking (31%). These services are often completed in concert with the elderly client. For example, a PCA may help the client bathe, cook, or clean. It is important to recognize that these supplemented ADLs are performed cooperatively between the caregiver and the client. Thus, technology for activity recognition should be prepared to recognize cooperative activities performed simultaneously by 2 or more individuals.

Monitoring use of assistive technology. A promising area for activity monitoring lies in collecting information about assistive equipment that may be in use. Interviewees and questionnaire respondents indicated that assistive equipment is common among clients, and that part of the CM's job is to determine what assistive equipment is necessary and then to arrange for that equipment to be provided. We hypothesize that activity recognition technology should monitor use of assistive technology, both to make sure that clients are safe and to help the CM make better decisions about what assistive technology is necessary (i.e., actually used).

Nutrition. Monitoring the nutrition and eating habits of clients is a promising area for activity recognition research. Every interviewee indicated that detecting malnutrition and improving eating habits was an important part of her job. One interviewee claimed, *"Their nutrition is big. If they don't eat...they seem to become confused, they get weaker, they're more fatigued, and they're higher risk for going into a nursing home."* 84% of questionnaire respondents identified eating and nutrition as very important to know, and we ranked it as the third "most useful" ADL related activity to know (see Table 2.3). When asked "If a magic genie could remind all of your clients to do one thing every day, what should the genie remind them of?," 10% of respondents said "eat better." 65% of respondents also said that food allergies were very common among their clients. Clearly, monitoring nutrition and eating habits is an important area for CMs and future technology.

Supplementing, not replacing, paperwork. Initially, it seems that automatic activity recognition could be used to automatically fill out paperwork. Recall that the most common response for both interviewees and questionnaire respondents to an open-ended question about what the most challenging aspect/hardest part of their job was completing paperwork. Additionally, respondents reported spending a mean of 70 minutes per day on paperwork. If technology could reduce this to 0 minutes, this would be around a 12% productivity boost.

However, CMs were not interested in technology to replace their job; rather, they expressed a need for more/better information so that they could improve the quality of their work. Every interviewee described visiting clients as their favorite part of the job ("[the home visit] *is the biggest part of my job that I enjoy. Usually it's the highlight of their* [the client's] *day to have the skilled nurse, case manager, or their caregivers come in."*), and many mentioned a desire that technology not replace face-to-face contact. In addition, 97% of the 91 questionnaire respondents indicated that they would still choose to visit their clients even if a magic genie was able to inform them of all the information necessary to fill out forms. This indicates that technology that is designed to help CMs should not replace work tasks, but should provide data to "fill in the gaps" in what CMs are able to observe, and/or to buttress their conclusions.

2.6.3 The Elder's Care Network

The physical and mental well-being of an elder is intimately tied to a "care network" – the group of people, representing both informal and formal support, involved in caring for the elder. Projects at Intel Research Seattle (e.g., Computer-Supported Coordinated Care (CSCC) and the CareNet Display) and elsewhere (e.g., Georgia

Tech's Digital Family Portrait) have been aimed at understanding the challenges for the care network and providing services that will assist, connect, and expand it [36, 35, 101, 83]. Results from this study support crucial design decisions – determining what information is valuable, what information is private, and who should have access to what information.

Sharing ADL information. Interviewees pointed out that elderly clients are not always receptive to sharing information, even with loved ones (*"If* [my daughter] *wants to see how I'm doing then she can call me.* [Note: the CM was describing a conversation]"). Questionnaire respondents also indicated that it is rare for a spouse (13%) or children (2%) to see the paperwork the CM fills out for the client. Interestingly, there seems to be little contact between doctors and CMs. One interviewee says, *"Every consumer* [client] *has to have a doctor. And you would think you'd have contact with the doctor but you really don't. Because the* [PROGRAM] *is really kind of non-medical."* Respondents also indicate rarely interacting with doctors face-to-face (only 14% report doing so), and instead, usually interact with them over the phone (93%). 32% of respondents indicated that doctors access client forms. Respondents indicate that they believe clients trust doctors the most (34%), followed by family members (25%). These findings indicate that future technology would benefit from efficient information sharing protocols that can connect formal support, informal support, and clients – while respecting client confidentiality requests.

Monitoring ADLs when there are communication challenges. We found that CMs depend on information from outside sources (i.e., members of the care network) when clients suffer from dementia or communication difficulties (e.g., language barriers or disabilities). Interviewees reported seeking information from informal support when clients have cognitive problems (*"Sometimes if you have someone that does have some dementia...you have to make sure to call their family."*) or disabilities (*"I've got one* [client] *who is like stone deaf. I'll call the daughter and just say, I'm going to see your mom. Does she need anything?"*). Questionnaire respondents also indicated asking informal support for help when unable to communicate with clients (35%). In addition, 10% of respondents indicated that client dementia/agitation makes activities difficult to monitor and 45% indicated occasionally using an interpreter. These findings indicate that communication challenges are a fact of life for case managers, and that future technology should be designed with this in mind.

Engaging support. A large portion of the job for a CM is to keep track of informal and formal support and services. Two interviewees mentioned connecting informal support to a client (i.e., ensuring that informal support is living up to agreements) as the hardest part of their job (*"Family says they're providing twenty-four hour care. They say that they're bringing food in. They say that they're helping them bathe on weekends. And then you find out that they haven't been over there for three weeks."*). This may be difficult, because some informal support does not want to help clients with their ADLs. One interviewee frankly stated, *"You don't want to give your mom a bath...or change her diapers."* In addition, interviewees mentioned that part of the CM's job is to ensure that formal support – specifically aides, or PCAs – are performing their jobs (*"If they are in the plan supposed to be doing something, I'm gonna be talking to em."*) and to make sure services are delivered and used (*"I make sure they have their safety equipment."*). Thus, CMs are in a position of both allocating resources (including the help of others) into a service plan, and then enforcing that plan. This is a difficult task and a prime area for future technology.

Educating the client about services. Although a CM may determine which services are necessary, it is up to the client to decide which services to accept. Thus, CMs repeatedly mention that encouraging clients and educating clients consumes a lot of time. One interviewee described this process as the hardest part of her job, *"The hardest part is, you know, individuals are legally able to make their own decisions. And so it's difficult working with people who make wrong decisions."* Another interviewee described the process as a finely honed job skill, *"Nobody wants to be wearing diapers, but if they understand they're like underwear..."* Yet another interviewee mentioned that uncooperative clients often require more time from the CM (and more visits), *"if they're adamant then we usually visit them more often."* 10% of questionnaire respondents indicate that non-compliant clients are the most difficult part of the job. Clients are often hesitant to accept services (an interviewee explained, *"They* [clients] *don't want to admit to being dependent on somebody to help em."*), and/or ignorant of the services that are available (*"A lot of times they* [the clients] *don't understand that the program is designed to assist them to stay in their home as long as they're safe. They think that if they're not able to do something then we're gonna put em in a nursing home."*). When asked what advice they would give clients, 10% of questionnaire respondents said they would urge clients to ask for help – and accept it. Encouraging and educating clients about goods and services are key responsibilities of a CM, although they may be difficult to quantify.

Medication monitoring. 96% of questionnaire respondents identified taking medication as very important to know about. When asked the open-ended question of what they would like to remind their clients of every day, the number one response was to remind clients to take medications properly (56%). Every interviewee from phase #1 identified making sure clients were taking their medications as an important part of the job. Interviewees mentioned several possible problems concerning medication, including clients who forget to take it (*"Elderly people don't think about taking their medication."*), clients who abuse it (*"I will fax over my visit to a physician if he requires it...if we feel like there's medication abuse."*), or medication errors (*"if we notice...that they are taking one medication and not complementing it with another."*). This problem is already well recognized, but our findings stress again that technology to monitor medication use would be very useful to CMs.

2.7 Conclusion

In this chapter, we described our methodology and the results of a two-phased formative study focused on case managers (CMs), the professional caregivers who are responsible for assessing Activities of Daily Living (ADLs) in elderly clients. We described phase #1 of the ADL study, in which we interviewed five case managers, and phase #2 of the ADL study, in which questionnaires were collected from 91 case managers. We discussed the results of both phases of the study, focusing on the findings that related most to automatic in-home health assessment technology. Some implications were discussed toward three areas relevant to the field of ubiquitous computing research: people tracking, activity recognition, and the elder's care network.

Chapter 3

Activity and Location Inference

We seek to provide the information that is vital for automatic health monitoring: identifying people, tracking people as they move, and knowing what activities people are engaged in. In this chapter, we introduce the simultaneous tracking and activity recognition (STAR) problem.[1] The key idea is that people tracking can be improved by activity recognition and vice versa. Location and activity are the *context* for one another and knowledge of one is highly predictive of the other. The algorithms we describe provide simultaneous room-level tracking and recognition of locomotion (which we loosely categorize as an activity), as well as recognition of more complex activities of daily living (ADLs).

3.1 Introduction

Automatic health monitoring necessarily occurs in a home environment where privacy, computational, and monetary constraints may be tight. We proceed from the "bottom-up," using predominantly anonymous, binary sensors that are minimally invasive, fast, and inexpensive. We call a sensor anonymous and binary when it can not directly identify people and at any given time it reports a value of one or zero. These sensors can be found in existing home security systems.

We describe a particle filter approach that uses information collected by many

[1] This chapter is a revised version of [117].

simple sensors. Particle filters offer a sample-based approximation of probability densities that are too difficult to solve in closed form (e.g., tracking multiple occupants in a home environment via several hundred anonymous, binary sensors). Particle filters are desirable because they can approximate a large range of probability distributions, focus resources on promising hypotheses, and the number of samples can be adjusted to accommodate available computational resources. We show that a particle filter approach with simple sensors can tell us which rooms are occupied, count the occupants in a room, identify the occupants, track occupant movements, recognize whether the occupants are moving or not, and recognize several activities of daily living.

This chapter is organized as follows: In the next section we review existing instrumented facilities and discuss the state-of-the-art in location estimation and activity recognition. Next, we describe the sensors we use and our rationale behind choosing simple sensors. Afterwards, we introduce our approach, including the details of our learner. The next three sections contain experimental results from simulations and real instrumented environments. Finally, we discuss our findings and conclude.

3.2 Related Work

Over the last several years much effort has been put into developing and employing a variety of sensors to solve key problems in the ubiquitous computing domain, including camera networks for people tracking [121, 23, 104], as well as cameras and microphones for activity recognition [33, 79]. Wearable sensors have been used for health monitoring [69], the facilitation of group interactions [50], and memory augmentation [97]. In this section we discuss these efforts in terms of automatic health monitoring, people tracking, and activity recognition.

3.2.1 Automatic Health Monitoring

People tracking and activity recognition experiments typically occur in a laboratory setting in a corporate or educational building [57, 23, 26]. Recently, there has been an increase in the number of stand-alone instrumented home environments. The Aware

Home project at Georgia Tech has built a house instrumented with ubiquitous computing technology for a variety of experiments [2]. The house has been fitted with a great variety of sensors with the goal of helping elderly adults live independently by providing memory augmentation, accident detection, and behavioral trend tracking. Researchers at MIT working on the House_n project have purchased a house and instrumented it with their own version of generic, simple sensors [54]. Currently, they deploy sensors for weeks at a time, collect sensor data as well as occupant labeled activity data, and then retrieve sensors for off-line activity recognition. Initial results show that for multiple instrumented houses clustered activity episodes correspond to data labeled by occupants. Researchers at the University of Florida have also instrumented a house with ultrasound localization and displays with the goal of providing timely and relevant information to residents [48]. Finally, the Neural Network House sensed appliance use and environmental changes to train neural networks to control levels of energy conservation and comfort [80]. These laboratories have explored an exciting variety of sensors to solve a variety of highly interrelated problems, mostly subsets of localization and activity recognition. Usually, these instrumented homes do not host long-term residents. Other research groups, including our own, have instrumented actual health care facilities for a variety of experiments [7, 12, 77]. The instrumented home and apartment used in our experiments are unique in that they use cheap, off-the-shelf sensors for simultaneous tracking and activity recognition. These instrumented facilities are valuable testbeds for a variety of algorithms and sensor configurations. Our experiments are intentionally constrained by the needs of elderly inhabitants. Rather than providing specific services, this research focuses on forming models of and recognizing behavior.

There has been some research into using binary sensors for automatic health monitoring. For several years a group of researchers at the Tokyo Medical and Dental University have been instrumenting homes with sensors such as motion detectors and contact switches to collect data for months at a time [84, 86, 85]. Although learning algorithms have not been applied, the raw data generated during these experiments was made available to physicians who were able to pick out patterns of activity by hand. Researchers at the Medical Automation Research Center (MARC) at the

University of Virginia have used an array of motion detectors and contact switches to attempt to detect activities of daily living (ADLs) [12]. They cluster sensor readings into rough groups based on room, duration, and time of day and demonstrate that many of the clusters do correspond to ADLs. Researchers have solidly identified the potential of simple sensors for automatic health monitoring.

3.2.2 People Tracking

People tracking is a fundamental problem in ubiquitous computing and has been approached via a variety of sensors, including cameras, laser range finders, wireless networks, RFID (Radio frequency identification) badges, and infrared or ultrasound badges [1, 3, 18, 33, 61, 80, 44, 104]. See [49] for a survey of location estimation techniques. A distributed network of many low cost sensors has several advantages over co-located sensors on a single platform (e.g., wearable sensors or mobile robots). The total coverage may be much larger and redundancy may exist between overlapping sensors. Also, sensor networks are more robust against failure or loss of individual components. In this research it was valuable to be able to quickly replace malfunctioning sensors, although sensor network robustness issues were not explored. We have chosen to explore a set of sensors that are already present in many homes as part of security systems. These sensors are cheap, computationally inexpensive, and do not have to be continuously worn or carried. We aim for room-level tracking, as our sensors do not provide the higher spatial resolution of other types of tracking systems.

Combining anonymous sensors and sensors that provide identification information for people or object tracking is an open problem. In the multi-target tracking community it is commonly known as the *data association* problem. The goal is to associate a set of current measurements with a set of existing "tracks" or object trajectories. In AI literature the problem of *object identification* is essentially the same, to determine if a newly observed object is the same as a previously observed object. A technique introduced by [51] uses pairwise sensor-based *appearance probabilities* to match images of cars between two traffic cameras. Researchers at Berkeley noted that this technique could not scale as more sensors were added. They used a Markov chain Monte Carlo approach to make the problem tractable for accurately tracking a single car between many cameras with minimal noise [89]. The vast number of possible *assignments* and noisy real-world data has spurred a variety of probabilistic approaches. Bayesian techniques, particularly particle filters, have been introduced as effective solutions [10, 62, 52]. In a recent experiment a particle filter implementation used laser range finders and infrared badges to track six people simultaneously in an office environment for 10 minutes [43]. The range finders provided anonymous, high granularity coordinates while the badge system identified occupants. We also use a particle filter approach to solve the data association problem, however, we use ID sensors only at entrances and exits and rely upon individual motion and activity models to resolve ambiguity within the environment. Data collected over the long-term provides an ever-improving model of the unique patterns of each occupant. We explore the ability of these models to identify occupants in lieu of additional ID-sensors.

3.2.3 Activity Recognition

An impressive amount of research falls under the umbrella of *activity recognition*. In particular, researchers have used cameras to detect a variety of activities, including sign language recognition [105], human gait recognition [76], sitting, standing and walking behaviors via wearable cameras [69], and recognizing American football [53] and basketball [59] plays from video. A variety of other sensors have been applied as well, including GPS readings to infer walking, driving and bus riding behaviors [91], laser range finders to learn motion paths in a home [18], audio to recognize conversational interactions over cell phones [9] and bathroom activities [29], and combinations of audio and video to recognize behavior in an office environment [87], group meeting interactions [31], and interactions between individuals [33]. Researchers at Intel Research have used radio frequency identification tags to recognize several ADLs [42]. Additionally, in recent work at the University of Washington, researchers have used location collected by GPS readings to help infer high-level activity such as shopping, working, or being at home [71]. However, we are unaware of any research that has attempted to use machine learning techniques to automatically recognize multiple ADLs using the sort of binary sensors common to home security systems, with or without simultaneous people tracking to improve recognition results.

A growing variety of Bayesian techniques have been used for activity recognition [87, 90]. For a survey of Bayesian techniques applied to activity recognition see [43]. In healthcare literature Dynamic Bayes nets have been used extensively for *execution monitoring*, a more intensive form of activity recognition in which the goal is to determine whether a person is following a plan appropriately [6, 106, 34]. Execution monitoring calls for recognition of specific parts of an activity as well as possible paths of progression through the plan. These approaches vary depending on the characteristics of the activity to be recognized. Our novel contribution arises from the interplay of tracking and activity recognition in an integrated system that uses only information from many anonymous, binary sensors.

3.3 Instrumenting the Home

In this section, we describe which sensors we use and why. First, we discuss our overall sensing goals for this chapter. Next, we discuss several challenges faced when placing sensors in a home and describe the ideal properties of sensors. Finally, we list the sensors used in these experiments.

3.3.1 Sensing Goals

Automatic health monitoring is predominantly composed of *location* and *activity* information. Below is a list of exactly what we wish to automatically recognize.

- **Presence.** Determine how many and which people are in the environment.
- **Individual Identification.** Determine the identity of each person.
- **Room-Level Tracking.** Determine the location of each person.
- **Locomotion.** Recognize whether an occupant is moving or sitting still (e.g., walking, wheelchair use, etc.,).

- **Activities of Daily Living.** Recognize eating/drinking, housework, and toileting [64].
- **Instrumented Activities of Daily Living.** Recognize cooking [68].
- **Extended Activities of Daily Living.** Recognize using a computer and using the telephone [99].

3.3.2 Sensor Constraints and Issues

In the previous chapter, we found that cost of sensors and sensor acceptance are pivotal issues, especially in the home. Participants overwhelmingly indicated that elders would be uncomfortable living with cameras and microphones. We found that people are often unwilling, forget, change clothes too often, or are not sufficiently clothed when at home to wear a badge, beacon, set of markers, or RF tag. In particular, elderly individuals are often very sensitive to small changes in environment [24], and a target population of institutionalized Alzheimer's patients frequently strip themselves of clothing, including any wearable sensors [25]. As a result, there is a great potential for simple sensors to 1) "fill in the blanks" when more complex sensors can not be used and 2) to reduce the number of complex (and possibly expensive) sensors that are necessary to solve a problem.

Like other researchers in academia and industry, we envision an off-the-shelf system installed and configured by a consumer [72, 12, 15, 108]. Ideally, the sensors we choose should offer solutions to the following issues: sensors and monitoring systems should be *invisible* or should fit into *familiar* forms. Sensor data should be *private* and should not reveal sensitive information, especially identity. Arguably equally important – sensors should not be perceived as invasive. Sensors should be *inexpensive*, preferably available off-the shelf. Sensors should be *easy to install.* Wireless sensors can be mounted to any surface, while wired sensors may require running cable to a central location. Processing sensor data should require *minimal computational resources* (e.g., a desktop computer). Sensors should be *low-maintenance*, easy to replace and maintain. Sensors will be neglected and should be robust to damage. Finally, sensors should be *low-power*, requiring no external power or able to run as

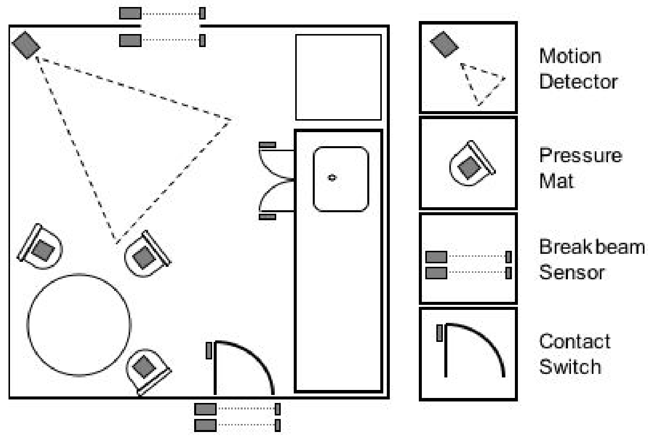

Figure 3.1: Overview of a typically instrumented room.

long as possible on batteries. As a last resort the device may need to be plugged in or powered by low voltage wiring.

3.3.3 Anonymous, Binary Sensors

Sensors that are *anonymous* and *binary* satisfy many of these properties. Anonymous sensors satisfy privacy constraints because they do not directly identify the person being sensed. Perceived privacy issues are minimized by the fact that anonymous, binary sensors are already present in many homes as part of security systems. Binary sensors, which simply report a value of zero or one at each time step, satisfy computational constraints. These sensors are valuable to the home security industry because they are cheap, easy to install, computationally inexpensive, require minimal maintenance and supervision, and do not have to be worn or carried. We choose them for the same reasons, and because they already exist in many of our target environments. (We typically use a denser installation of sensors than in a home security system, however.)

Sensor Choice and Placement

For our experiments, we chose to use commonly available anonymous, binary sensors, including: motion detectors, break-beam sensors, pressure mats, contact switches, water flow sensors, electric current sensors, and wireless object movement sensors. Motion detectors are placed near the ceiling in order to maximize room coverage. Contact switches are placed on doors and drawers of all types, including cabinets and refrigerators. Pressure mats are placed under couches, chairs and rugs. Break-beam sensors are triggered by occupants walking through the beam. With two beams we can infer direction. Water flow sensors are placed on hot and cold water pipes and toilets. Current sensors are placed near transformers, to monitor the amount of current flowing to circuits, e.g., electrical outlets. Wireless object movement sensors are placed on a variety of objects to detect when those objects are manipulated. See Figure 3.1 for an overview of a typically instrumented room.

In addition, we solve the presence and identity problems by using an ID sensor to capture identity as occupants enter and leave the environment. In experiment # 2, we replaced house keys with unique radio frequency identification (RFID) tags. Instead of a lock and key, an RFID reader near the doorway "listens" for the key (an RFID tag) and automatically records identification while it unlocks the door for a few seconds. (The RFID reader can detect multiple keys simultaneously from a distance of about a foot.) Afterwards, the door locks itself and the occupant need not continue to carry the key. Currently, we use the following sensors:

- **Motion detectors.** Motion detectors provide a binary indication of heat and movement (e.g., human presence) in an area. In experiment # 2 we used X10 Hawkeye motion detectors. After each reading these sensors pause for eight seconds before becoming active again.The detectors are wireless, pet-resistant, require heat and movement to trigger, and run on battery power for over one year. In experiment # 3 we used motion information reported by cameras. The cameras compared each new frame to the previous frame and reported movement when the difference between frames exceeded some threshold.

- **Break-beam sensors.** We use these sensors in groups of two to determine when an occupant passes through a doorway and in what direction. They work by generating a beam across a space and monitoring when it is reflected back. While the beam is interrupted the sensor changes state.

- **Pressure mats.** These wired sensors were used to detect presence on chairs and couches. The pressure mats are made of two metal screens separated by a piece of foam with holes. The weight necessary for contact depends on the size and number of holes cut into the foam layer.

- **Contact switches.** These inexpensive wired magnetic contact switches indicate a closed or open status. They are attached to all manner of doors, drawers, and cabinets.

- **Water flow sensors.** When placed into water pipes these sensors report a reading when flow exceeds some threshold.

- **Current sensors.** These sensors measure current flow in a circuit, reporting when current exceeds some threshold, e.g., whenever an appliance is used.

- **Wireless Object Movement Sensors.** These small battery-powered sensors, called MITes, are designed to be attached to movable physical objects. They consist of a single two-axis 2G accelerometer and an RF transceiver. The MITes work by measuring acceleration and broadcasting a unique ID whenever movement exceeds a sensitivity threshold. For a detailed description of MITe hardware, see [56].

- **Radio Frequency Identification (RFID).** We use low frequency RFID to identify occupants entering and leaving the environment. The system sends a modulated RF signal to an antenna, which amplifies the signal, creating a small field near the front door. When the credit card-sized transponder or 'tag' is in the field, an integrated circuit detects the signal and uses its energy to send a unique identification signal. This signal is decoded and sent to a computer via an RS-232 interface. The entire process takes less than 100ms and multiple tags can be read simultaneously. Each occupant is given a unique tag; upon recognition the tag will automatically unlock the door, as well as identify the occupant entering or leaving the environment. This interface faces challenges when occupants forget badges or when guests visit.

A Note on Privacy

Our decision to use simple sensors provides inherent privacy at the physical layer, but does not directly address higher-level privacy issues, such as dissemination of information. Aside from the RFID antenna (which requires a tag), none of the individual sensors in this research can be used to identify a person. The locations and activity information collected by applying machine learning algorithms to the entire collection of sensor information is obviously of a private nature. The dissemination of and access privileges to this information (whether to family, physicians, or to the general public) will depend on the services provided using this system, and are outside of the scope of this document.

3.4 Approach

In this section, we introduce the STAR problem, discuss why it is difficult to solve with simple sensors, and consider several simplifications. We discuss a Bayes filter approach and show why it fails to accomodate multiple occupants. We then describe a Rao-Blackwellised particle filter that is able to handle multiple occupants by performing efficient data association. We discuss how to learn model parameters both online and offline.

3.4.1 Simultaneous Tracking & Activity Recognition

There are two main problems when solving STAR for multiple people, (1) what is the state of each person and, (2) which person is which? In the first problem, observations are used to update the state of each occupant (i.e., their activity and location). In the second problem, identity of the occupants is estimated and anonymous observations are assigned to the occupants most likely to have generated them. Uncertainty occurs when several occupants trigger the same set of anonymous sensors. The tracker does not know which occupant triggered which sensor (i.e., which data to associate with which occupant).

There are several ways to simplify the problem. First, we could *increase the number of ID sensors.* This simple approach solves the problem by using sensors that identify occupants outright. Unfortunately, ID sensors are expensive, have significant infrastructure requirements, and/or must be worn or carried by the occupant. It is more desirable to employ many inexpensive sensors in lieu of expensive sensors. Second, we could *increase the sensor granularity.* Adding more sensors can reduce the ambiguity caused by multiple occupants, but may be expensive. Alternately, existing sensors can be placed so that they collect the maximum amount of information. In experiments, we intentionally placed sensors so that they would detect different properties, which increases granularity. For example, ceiling-mounted motion detectors detect gross movement while chair-mounted pressure mats detect static occupants. Similarly, noting which contact switches are out of reach of pressure mats can potentially separate two occupants when one is seated and the other opens a drawer. Third, we could *learn individual movement and activity patterns.* Over time, statistical models can represent particular habits of select individuals. Individualized motion models can help the tracker recover from ambiguity as occupants follow their normal routines (e.g., sleeping in their own beds).

3.4.2 Bayes Filter Approach

First, we address the question of how to update occupant state given sensor measurements. Bayes' filters offer a well-known way to estimate the state of a dynamic system from noisy sensor data in real world domains [40]. The *state* represents occupant location and activity, while sensors provide information about the state. A probability distribution, called the *belief*, describes the probability that the occupant is in each state $p(X_t = x_t)$. A Bayes filter updates the belief at each time step, conditioned on the data. Modeling systems over time is made tractable by the Markov assumption that the current state depends only on the previous state.

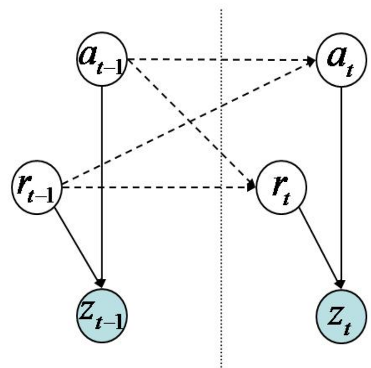

Figure 3.2: A DBN describing room-level tracking and activity recognition.

We estimate the state $x_t = \{x_{1t}, x_{2t}, ..., x_{Mt}\}$ of M occupants at time t using the sensor measurements collected so far, $z_{1:t}$. At each time step we receive the status of many binary sensors. The measurement $z_t = \{e_{1t}, e_{2t}, ..., e_{Et}\}$ is a string of E binary digits representing which sensors have triggered during time step t. The update equation is analogous to the forward portion of the forward-backward algorithm used in hidden Markov models (HMMs). See [96] for a detailed description of how HMMs work.

$$p(X_t = x_t|z_{1:t}) \propto p(z_t|X_t = x_t) \sum_{x' \in X} p(X_t = x_t|X_{t-1} = x')p(X_{t-1} = x'|z_{1:t-1}). \quad (3.1)$$

The *sensor model* $p(z_t|X_t = x_t)$ represents the likelihood of measurement z_t occurring from state x_t. The *motion model* $p(X_t = x_t|X_{t-1} = x')$ predicts the likelihood of transition from the state x' to the current state x_t. How these models are learned is discussed in section 3.4.4.

The graphical model in Figure 3.2 represents the dependencies we are about to describe. The state space $x \in X$ for occupant m is the range of possible locations and activities, $x_{mt} = \{r_{mt}, a_{mt}\}$, where $r \in R$ denotes which **r**oom the occupant is in, and $a \in$ {moving,not moving} denotes occupant **a**ctivity. The raw sensor values are the only given information; the rest must be inferred. Each observation is composed of a collection of *events* and appear $z_t = \{e_{1t}, e_{2t}, ..., e_{Et}\}$. Event generation is straightforward. For example, an event is generated when a motion detector triggers, when a contact switch changes state, or when a wireless object movement sensor is moved.

Tracking multiple people causes the state to have quite large dimensionality, making model learning intractable. Currently, a simplifying independence assumption between m occupants means that the update equation is factored as:

$$p(X_t = x_t | X_{t-1} = x') = \prod_{m \in M} p(X_{mt} = x_{mt} | X_{m,t-1} = x'_m). \tag{3.2}$$

This assumption could be partially relaxed through the use of two models, one for occupants that are alone and another for multiple occupants. This abstraction avoids the exponential blow up resulting from joint models of combinations of specific individuals. A similar approach has been applied successfully to tracking multiple interacting ants in [66].

Equation 3.1 describes the Bayes filter update using all observations up to the current time step $z_{1:t}$. Higher accuracy is usually obtained off-line by using past and future information at each time step. This is commonly known as *smoothing*. Smoothing provides higher accuracy for off-line purposes, such as a daily summary of activity [116]. The update equation is analogous to the *backward* step of the forward-backward algorithm commonly used in HMMs. We report results using smoothing in experiment # 3.

$$p(X_t = x_t | z_{t+1:T}) \propto \sum_{x \in X} p(X_{t+1} = x | z_{t+2:T}) p(X_{t+1} = x | X_t = x_t) p(z_{t+1} | X_{t+1} = x). \tag{3.3}$$

Classical Data Association Methods

The above approach works well for tracking a single occupant in a noisy domain (the Bayes filter is named for its ability to *filter* spurious noise). However, this approach struggles to track multiple occupants because other occupants do not behave like noise processes. The tracker becomes confused by constantly conflicting sensor measurements. We need some way to determine which occupant generated what observation. This is the data association problem, and in our domain it can become severe. For t seconds and m occupants each association has $m!^t$ possibilities. In a reasonable scenario with several dozen inexpensive sensors monitoring a handful of occupants for a week, there are too many data assignments to enumerate.

There are several classical data association methods (for a survey see the book [11]). Probably the simplest approach, called the *nearest neighbor standard filter* (NNSF), uses only the closest observations to any given state to perform the measurement update step. This method has a hard time recovering lost targets because unlikely observations are ignored. A more accurate method is called the *probability data association filter* (PDAF), which uses the probability of an observation from a target versus from clutter to assign weighted measurements. This approach fails for the same reason independent Bayes filters do – occupants do not behave like noise. This problem is dealt with by the *joint probability data association filter* (JPDAF), which finds the joint probability of all possible assignments for the current time step. JPDAF then updates the state by a sum over all the association hypotheses weighted by the probabilities from the likelihood. An even more general method, called *multi-hypothesis tracking* (MHT), calculates every possible association hypothesis over time as well. PDAF, JPDAF, and MHT approaches require exhaustive enumeration of every possible association, which can quickly become intractable. There are many work-arounds, including gating and pruning trees of possible hypotheses. Recently, particle filters have been applied successfully to the data association problem [19].

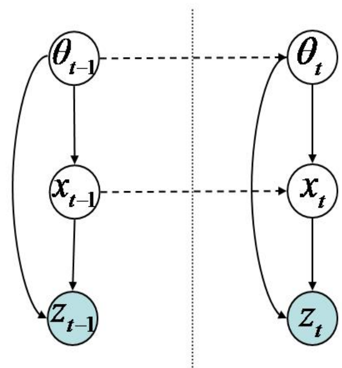

Figure 3.3: A DBN describing occupant state and data associations.

3.4.3 Particle Filter Approach

At each time step we wish to find the best assignment of sensors to occupants and to use this assignment to update the state of each occupant. Assignments between sensor measurements and occupants are not given. Therefore, we must now estimate the posterior distribution over both occupant state and sensor assignments.

We let θ_t represent a sensor assignment matrix such that $\theta_t(i, j)$ is 1 if event e_{it} belongs to occupant j and 0 otherwise. See Figure 3.3 for the updated graphical model. We must expand the posterior of Equation 3.1 to incorporate data association. We accommodate our expanded posterior efficiently by using a Rao-Blackwellised particle filter [40]. By the chain rule of probability,

$$p(X_{1:t}, \theta_{1:t}|z_{1:t}) = p(X_{1:t}|\theta_{1:t}, z_{1:t})p(\theta_{1:t}|z_{1:t}). \tag{3.4}$$

The key idea is to update the *state* $p(X_t = x|\theta_{1:t}, z_{1:t})$ analytically using the Bayes filter update already described, and to use a particle filter to generate a sample-based approximation of *assignments* $p(\theta_{1:t}|z_{1:t})$. This streamlines our approach by sampling only from the intractable number of possible sensor assignments and solving exactly for our (relatively) small number of possible state configurations.

The desired posterior from Equation 3.4 is represented by a set of N weighted particles. Each particle j maintains the current state of all occupants via a bank of

M Bayes filters, as well as the sensor assignments and the weight of the particle.

$$s_t^j = \{x_t^{(j)}, \theta_{1:t}^{(j)}, w_t^{(j)}\}. \tag{3.5}$$

Note that for filtering purposes we may store only the latest association $\theta_t^{(j)}$. $x_t^{(j)}$ is a distribution over all possible states of all occupants. The $\theta_t^{(j)}$ are updated via particle filtering, and the $x_t^{(j)}$ are updated exactly using the Bayes filter update. The marginal distribution of the assignment (from Equation 3.4) is therefore approximated via a collection of N weighted particles,

$$p(\theta_{1:t}|z_{1:t}) \approx \sum_{j=1}^{N} w_t^{(j)} \delta(\theta_{1:t}^{(j)}, \theta_{1:t}). \tag{3.6}$$

where $w_t^{(j)}$ is the *importance weight* of particle j, and $\delta(x, y) = 1$ if $x = y$ and 0 otherwise.

Given the sample-based representation of assignments from Equation 3.6, the marginal of the state node is,

$$\begin{aligned}
p(X_t|z_{1:t}) &= \sum_{\theta_{1:t}} p(X_t|\theta_{1:t}, z_{1:t}) p(\theta_{1:t}|z_{1:t}) && (3.7)\\
&\approx \sum_{\theta_{1:t}} p(X_t|\theta_{1:t}, z_{1:t}) \sum_{j=1}^{N} w_t^{(j)} \delta(\theta_{1:t}^{(j)}, \theta_{1:t}) && (3.8)\\
&= \sum_{j=1}^{N} w_t^{(j)} p(X_t|\theta_{1:t}^{(j)}, z_{1:t}). && (3.9)
\end{aligned}$$

Given a sampled data association $\theta_{1:t}^{(j)}$ and an observation z_t, it is straightforward to update the belief $p(X_t = x|z_{1:t}, \theta_{1:t})$ exactly according to a slightly modified version of the Bayes filter from Equation 3.1. First, we show the predictive distribution, where information up to time step $t-1$ is used to predict the next state for particle j.

$$p(X_t = x | z_{1:t-1}, \theta^{(j)}_{1:t-1}) = \sum_{x'} p(X_t = x | X_{t-1} = x') p(X_{t-1} = x' | z_{1:t-1}, \theta^{(j)}_{1:t-1}). \quad (3.10)$$

We derive the full update equation given information up to time t according to Bayes rule.

$$\begin{aligned} p(X_t = x | z_{1:t}, \theta^{(j)}_{1:t}) &= \frac{p(z_t | X_t = x, \theta^{(j)}_t) p(X_t = x | z_{1:t-1}, \theta^{(j)}_{1:t-1})}{\sum_x p(z_t | X_t = x, \theta^{(j)}_t) p(X_t = x | z_{1:t-1}, \theta^{(j)}_{1:t})} & (3.11) \\ &\propto p(z_t | X_t = x, \theta^{(j)}_t) p(X_t = x | z_{1:t-1}, \theta^{(j)}_{1:t-1}). & (3.12) \end{aligned}$$

Given these definitions we now discuss the overall approach. The following sampling scheme, called *sequential importance sampling with re-sampling*, is repeated N times at each time step to generate a full sample set S_t (composed of samples $s^{(j)}_t$ where $j = 1...N$) [40].

During *initialization* occupant location and identity are gathered by RFID and sensor measurements are assigned automatically. In each iteration there are four steps. First, during *re-sampling* we use the sample set from the previous time step S_{t-1} to draw with replacement a random sample $s^{(j)}_{t-1}$ according to the discrete distribution of the importance weights $w^{(j)}_{t-1}$. Next, we *sample* a possible sensor assignment matrix $\theta^{(j)}_t$. We discuss how to propose sensor assignments in the next section. Next, we use the association $\theta^{(j)}_t$ to perform an *analytical update* of the state of each occupant in sample j via Equation 3.11. Finally, during *importance sampling* we weight the new sample $s^{(j)}_t$ proportional to the likelihood of the resulting posteriors of the state of each occupant. This is equal to the denominator of Equation 3.11,

$$w^{(j)}_t = \eta \sum_x p(z_t | X_t = x, \theta^{(j)}_t) p(X_t = x | z_{1:t-1}, \theta^{(j)}_{1:t}), \quad (3.13)$$

where η is a normalizing constant so that the weights sum to one.

The Data Association Problem

During the sampling step a possible assignment of sensor readings to occupants (a data association) must be proposed for the new sample. Choosing an impossible association will cause that particle to have a zero weight and wastes computational time. For example, foolishly assigning two sensors from different rooms to the same occupant will result in a particle with negligible probability. A more efficient particle filter will propose data associations in areas of high likelihood. The better the proposals, the fewer particles necessary.

Assigning sensor readings uniformly (regardless of occupant state) is inefficient because it will propose many unlikely or impossible associations (e.g., one occupant given sensor readings from different rooms). A quick improvement is to use *gating* to eliminate impossible associations, but a gated uniform method is still inefficient because it ignores the current state of each occupant. Sensors are intimately tied to rooms and activities. Occupants that are in the same room as a sensor are more likely to have triggered it and occupants engaged in certain activities are more likely to trigger associated sensors. A simple heuristic takes advantage of these properties. We currently assign measurements based on the posterior $p(\theta_t|x_{t-1}^{(j)})$. The proposed assignment matrix θ_t is constructed by independently assigning each measurement to an occupant based on the probability that she triggered it $p(e_{it}|x_t)\forall i$. This method tends to choose likely assignments and usually avoids impossible assignments, but is not guaranteed to approximate the true distribution $p(\theta_t|z_{1:t})$.

3.4.4 Parameter Learning

Modeling the behavior of individual occupants can increase tracking and activity recognition accuracy and make data association more efficient. In a system with few ID sensors (like ours) these models are vital to disambiguate the identities of many occupants. Motion models describe individual tendencies to transition between rooms and activities. Sensor models describe individual tendencies to set off specific sensors (e.g., shorter occupants may use high cabinet doors less often). Models can be initialized generically for unknown occupants.

Motion model. We wish to learn individual parameters for the motion model.

$$p(X_t = x_t | X_{t-1} = x_{t-1}) = p(a_t, r_t | a_{t-1}, r_{t-1}) \tag{3.14}$$
$$= p(a_t | a_{t-1}, r_{t-1}) p(r_t | r_{t-1}, a_{t-1}). \tag{3.15}$$

- $p(r_t | r_{t-1}, a_{t-1})$ is the probability of transition to a room given the previous room and occupant activity. Transition probabilities between contiguous rooms are initialized uniformly for active occupants and set to small values for idle occupants.

- $p(a_t | a_{t-1}, r_{t-1})$ models the probability of which activity the occupant is engaged in given the previous room and what the occupant's activity was during the last time step. This is initialized so that it is more likely for active occupants to continue to be active and idle ones to continue not to.

Sensor model. Individual sensor readings, called ***events***, are considered independent. For occupant m the sensor model can be rewritten:

$$p(z_t | X_t = x_t, \theta_t^{(j)}) = \prod_{m \in M} p(z_t | X_{mt} = x_{mt}, \theta_t^{(j)}) = \prod_{m \in M} \prod_i p(e_{it} | X_{mt} = x_{mt}, \theta_t^{(j)}). \tag{3.16}$$

This models the probability of observing each sensor measurement given the location and activity of the occupant. This sensor model is initialized by assigning small probability to sensor readings occurring outside their designated room.

This system uses a non-metric, room based location representation and a discrete set of mutually exclusive activities. The result is a relatively small number of discrete states, even when confounded with additional activities. This simplicity helps

Model Initialization:
1. Initialize model parameters with generic values.

E-step:
1. Generate N samples uniformly.
2. Forward filtering : for $t = 2...T$

(a) Generate N samples using the samples from the previous time step.
(b) Reweight each sample based on current observation z_t.
(c) Multiply or discard samples based on their weights.
(d) For each occupant m count and store $\alpha_t^m(r_t, a_t)$

3. Generate N samples uniformly.
4. Backward filtering : for $t = T..1$

(a) Calculate backward parameters $p(r_{t-1}|r_t, a_t), p(a_{t-1}|a_t, r_t)$
(b) Generate N samples using the samples from existing samples using backward parameter estimation.
(c) Reweight each sample based on current observation z_t.
(d) Multiply or discard samples based on their weights.
(e) For each occupant m count and store $\beta_t^m(r_t, a_t)$.

M-step:
1. Calculate γ_t^m and δ_t^m using equations (3.17) and (3.18) and then normalize.
2. Update parameters using equations (3.19) and (3.20).

Repeat

Table 3.1: Monte Carlo EM approach.

make *unsupervised* learning of model parameters possible. It also invites an intuitive understanding of how transitions occur between rooms and activities.

Training model parameters is simple when we know the true state of each occupant. The simplest approach is to train parameters on data generated by occupants that were home alone. While a person is home alone we can assume that any sensor readings are generated by that person or a noise process. Parameter learning can be performed with simple counting. This method ignores a significant amount of training data because occupants are often home together. It also fails to learn the difference between how people behave alone versus in the presence of others.

Multiple occupants introduce uncertainty that could hurt the accuracy of learned models. A common method to minimize this uncertainty is to use the Expectation-Maximization (EM) algorithm [20]. The EM algorithm is an iterative approach to finding parameters that maximize a posterior density. The idea is to use current model parameters to estimate the expectations (E-step) of the distribution. The model parameters are then updated (M-step) using the expectations from the E-step. The steps are repeated and in each iteration the model parameters are improved. Eventually the algorithm converges to a local maximum.

A version of the EM algorithm called Monte Carlo EM [70, 115] takes advantage of the set of particles representing the posterior. Researchers at Intel used this technique with GPS readings to learn models of movement and transportation methods of a traveler in the city [91] . In this version both forward and backward updates are applied to the Bayes filter at each time step. At each forward and backward step, the algorithm examines each particle and counts the number of transitions between rooms and activities for each occupant. The counts from forward and backward phases are normalized and then multiplied and used to update model parameters. The learning algorithm is introduced thoroughly for Monte Carlo HMMs in [109].

$\alpha_t^m(r_t, a_t)$ is the number of particles in which occupant m is in room r and performing activity a during the *forward* pass.

$\beta_t^m(r_t, a_t)$ is the number of particles in which occupant m is in room r and performing activity a during the *backwards* pass.

$\gamma_{t-1}^m(r_t, r_{t-1}, a_{t-1})$ is the probability that occupant m will move from room r_{t-1} to room r_t in activity a_{t-1} at time $t-1$.

$\delta^m_{t-1}(a_t, a_{t-1}, r_{t-1})$ is the probability that occupant m will change from activity a_{t-1} to activity a_t from room r_{t-1} at time step $t-1$.

We define, [109]

$$\gamma^m_{t-1}(r_t, r_{t-1}, a_{t-1}) \propto \alpha^m_{t-1}(r_{t-1}, a_{t-1}) p(r^m_t | r^m_{t-1}, a^m_{t-1}) \beta^m_t(r_t, a_{t-1}) \tag{3.17}$$

and

$$\delta^m_{t-1}(a_t, a_{t-1}, r_{t-1}) \propto \alpha^m_{t-1}(r_{t-1}, a_{t-1}) p(a^m_t | a^m_{t-1}, r^m_{t-1}) \beta^m_t(r_{t-1}, a_t) \tag{3.18}$$

After the counting phase we update parameters as:

$$p(r^m_t | r^m_{t-1}, a^m_{t-1}) = \frac{\sum_{t=2}^{T} \gamma^m_{t-1}(r_t, r_{t-1}, a_{t-1})}{\sum_{t=2}^{T} \sum_{r_t \in contiguous r_{t-1}} \gamma^m_{t-1}(r_t, r_{t-1}, a_{t-1})} \tag{3.19}$$

and

$$p(a^m_t | a^m_{t-1}, r^m_{t-1}) = \frac{\sum_{t=2}^{T} \delta^m_{t-1}(a_t, a_{t-1}, r_{t-1})}{\sum_{t=2}^{T} \sum_{a_t \in A} \delta^m_{t-1}(a_t, a_{t-1}, r_{t-1})} \tag{3.20}$$

See Table 3.1 for a summary of particle filtering with MCEM model learning.

Adaptation

A conceptually simple modification greatly increases the efficiency of our particle filter. The inefficiency arises during the re-sampling step, when samples from the previous sample set are drawn blindly, without considering the most recent observation. The particle filter is less wasteful with samples if the proposal distribution relies not only on the motion model, but also on the most recent measurement. This improvement is known as *assignment lookahead* or *adaptation* [95].

The goal is to update the sample weights of the previous timestep by the sample's ability to predict the observation of the current timestep. This update is applied to every particle $j = 1 \ldots N$ as soon as a new measurement z_t is received, and before the resampling step. For sample j at time step $t-1$ we wish to update by,

$$\varpi_{t-1}^{(j)} \propto w_{t-1}^{(j)} p(z_t | x_{t-1}^{(j)}) \tag{3.21}$$

Equation 3.21 is intractable because $p(z_t|x_{t-1}^{(j)})$ requires a summation over every possible data association. There are too many data associations to enumerate. We can approximate by using the previously described MCMC technique to approximate $p(z_t|x_{t-1}^{(j)})$ with a set of R samples $s_{t-1}^{(j)}$ where $j = 1...R$ and the sampled data associations are denoted $\theta_{rt}^{(j)}$. We use these sample associations to re-weight each member of the previous sample set.

$$\varpi_{t-1}^{j} \propto \frac{w_{t-1}^{j}}{R} \sum_{r=1..R} p(z_t|\theta_{rt}^{(j)}, x_{t-1}^{(j)}) p(\theta_t^{(j)}|x_{t-1}^{(j)}) \tag{3.22}$$

3.5 Experiment # 1: Simulated Data

In this section, we evaluate the performance of our approach on a simulated data set. The purpose of this experiment was to determine the feasibility of our approach. We are also interested in gauging the impact of unique motion models, sensor density, and parameter learning (online vs. offline) on tracking accuracy for multiple occupants.

3.5.1 Study Methodology

We implemented a simple program to simulate the data generated by occupants in an instrumented environment.[2] The simulator can generate data from any number of motion detector, contact switch, and pressure mats per room, as well as break beam sensors on doors between rooms. The number of occupants, room structure, doorway location, and noise rates can be specified via command line parameters. “Noise” is defined as a random sensor measurement. Each occupant obeys an independent first-order HMM motion model that is set by hand or initialized randomly. Sensors also obey a hand-set sensor model in which the likelihood that a given sensor will trigger

[2]The simulator can be downloaded from *www.danielhwilson.com*

Experiment	Accuracy
same	0.99 ± 0.0001
opposite	0.99 ± 0.0003
middle	0.66 ± 0.002
uniform	0.46 ± 0.01

Table 3.2: Comparison of motion model experiments.

depends upon the number of occupants in the room and whether they are moving or not.

Simulated occupants are introduced to the environment from the same starting state and identified correctly from this state, to imitate an RFID set up in the entry way. Henceforth, each occupant is unlikely to re-enter this state. The simulation differs from reality in that simulated occupants behave truly independently. Simulated occupants were active (moving) approximately 15% of the time. There was a sporadic sensor reading about once every ten minutes. The particle filter tracker used the same sensor model for each occupant. Parameters of motion models were either learned offline via counting, or online (i.e., during the experiment) via the EM Monte Carlo method.

Location and activity predictions are updated every second and accuracy is measured as the number of seconds in which the maximum likelihood predictions of the tracker match the labeled location tag (in simulated experiments the state of each occupant is known). Results are reported for real-time, online tracker performance.

3.5.2 Results

Motion model comparison. This experiment explored the value of unique motion models for tracking multiple occupants. We used six rooms, one entry way connected to a circular hallway composed of the other five rooms. Each room contained one motion detector, contact switch, and pressure mat. There were break beam sensors in every doorway. We tracked two simulated occupants for one hour with ten trials for each experiment. Occupants are identified in the entry room. They leave the entry room on the first time step and do not return.

We used four different motion models to generate the data and allowed the tracker to use the correct models. We ran four experiments (corresponding to the four motion models) called *same*, *opposite*, *middle*, and *uniform*. In the *same* experiment both occupants always walk in a clockwise direction. In the *opposite* experiment one occupant always walks clockwise and the other counter-clockwise. In the *middle* experiment, each occupant was %75 likely to transition clockwise. In the uniform experiment both occupants use the same uniform model in which they are equally likely to transition to any contiguous room. See Appendix D for the exact motion models.

We found that tracker accuracy depends heavily on occupant predictability. Accuracy was perfect in the *same* and *opposite* experiments, where each occupant performed the same action at each time step. Whether both occupants performed the same action (walking clockwise) or opposite actions (one clockwise and the other counter-clockwise) did not matter. Accuracy suffered in the *middle* experiment, when the transition probability was lowered to 75%. Accuracy was lowest for *uniform* models in which movements are completely unpredictable and identical between occupants. Results are summarized in Table 3.2.

Small house experiments. These experiments simulated a small house with ten rooms (three bedrooms, two bathrooms, a kitchen, living room, dining room, and hallways). Motion models for five occupants describe typical movements, with the first three occupants having their own bedrooms and the last two occupants as guests. Each experiment tracked occupants for one hour and was run for ten trials. In Figure 3.4, 3.5, and 3.6 the variance bars reflect variations over the ten trials.

Sensor configurations. First, we looked at the impact of sensor configurations on tracking accuracy (see Figure 3.4). In this experiment we tracked two occupants with generic motion models, using three different sensor configurations: the *normal* configuration contains one motion detector, contact switch, and pressure mat per room, the *extra* configuration contains three of each type per room (i.e., more chairs and cabinets were added to the room), and the *fewer* configuration contained only one motion detector per room. In general, more sensors improve accuracy. The *fewer* configuration had so few sensors that the number of particles ceased to matter. Also,

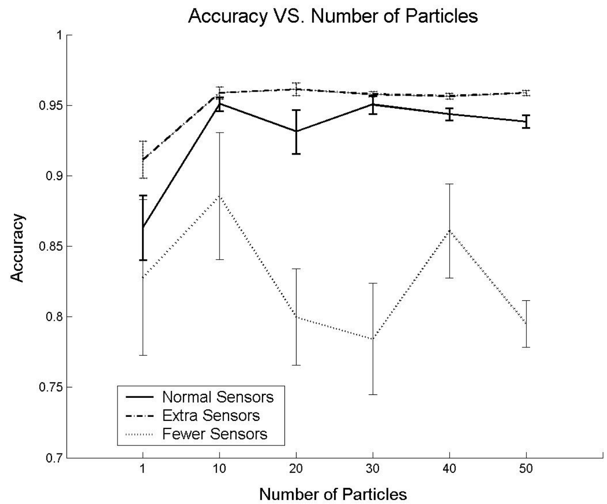

Figure 3.4: Accuracy vs. number of particles.

with fewer sensors come fewer measurements, and more variance on the periods before tracker recovery. Thus, the jagged line for the fewer configuration is caused by high variance between trials.

The number of particles will need to grow for sensor configurations with hundreds of sensors per room, which will pose a much more complex data association problem.

Parameter learning. Second, we examined how different approaches to model learning affect accuracy (see Figure 3.5). In this experiment, the number of particles is set to fifty and we compare three techniques for learning motion model parameters. One method is to use simple counting to train a model using data from when the occupant is home alone. Alternately, we can use probabilistic methods to train a model online, while several occupants may be home. Three methods were used to train model parameters: (1) learning motion models off-line given one day of data generated by occupants that are alone (*offline*), (2) on-line via the Monte Carlo EM algorithm (*online*), and (3) a combination in which the MCEM online parameter learning algorithm was seeded by a model already trained offline on one hour of single occupant data (*both*). In general, the *offline* method had highest accuracy, followed by *both* and with *online* learning last. Although the offline method performed best,

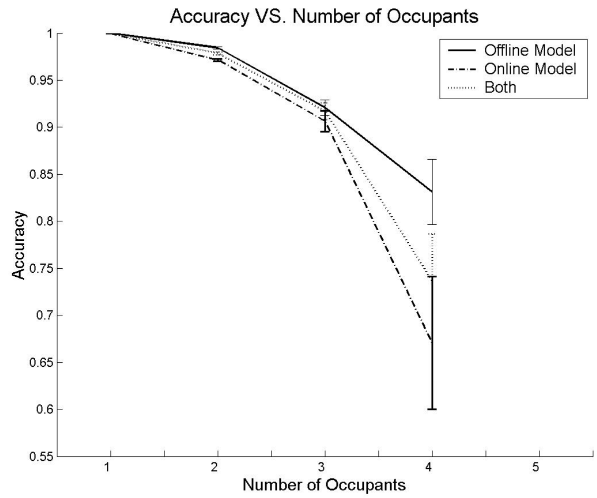

Figure 3.5: Accuracy vs. number of occupants.

this is due in part to the simplicity of our simulator, in which occupants behave independently. We feel that the *both* method, of seeding a model with offline data and continuing to learn online, is the most promising real-world approach. As the number of occupants rises from two to three to four, we see the *online* method take a big accuracy hit. This is expected, as online model learning will be confounded by multiple interfering occupants.

Number of occupants. In Figure 3.5 we varied the number of occupants and used fifty particles, and in Figure 3.6 we varied the number of particles and used offline model learning. Accuracy plateaus as the number of particles are increased. As the number of occupants increases the step from one to ten particles is increasingly important. Due to efficient data association methods, the tracker does not need hundreds of particles. Accuracy does not drop linearly as more occupants are tracked simultaneously; the difference between one and two occupants is much less than the difference between three and four occupants.

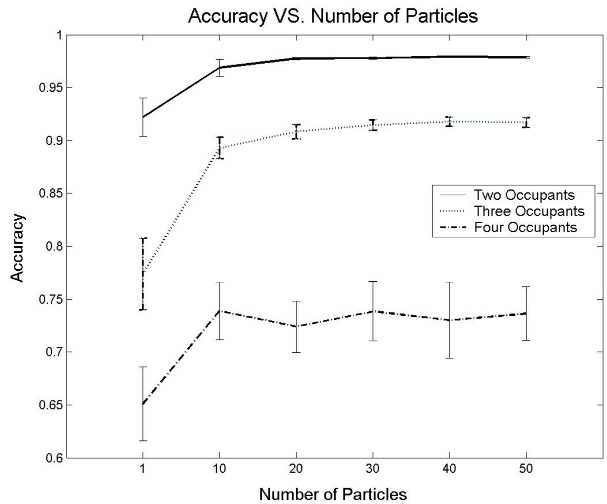

Figure 3.6: Accuracy vs. number of particles.

3.6 Experiment # 2: Instrumented House

In this section, we describe the methodology and reults of experiment # 2, in which we conducted experiments using data generated by one to three occupants in a real instrumented house. The purpose of this experiment was to verify that the particle filter approach could function in a real environment with significant noise factors, including misfiring sensors, broken sensors, and the presence of several pets.

3.6.1 Study Methodology

Participants. The participants in this experiment were three occupants who were permanent residents of the instrumented home. Participants included two males (including the author), one female, one dog, and one cat.

Instrumented Environment. The instrumented three story house was 2824 square feet and contained twenty separate rooms. The house contained one RFID reader located in the front doorway (the back door was not used during the experiment). There were twenty four motion detectors, with at least one per room. Twenty four contact switches were distributed to every doorway, the refrigerator, and in many

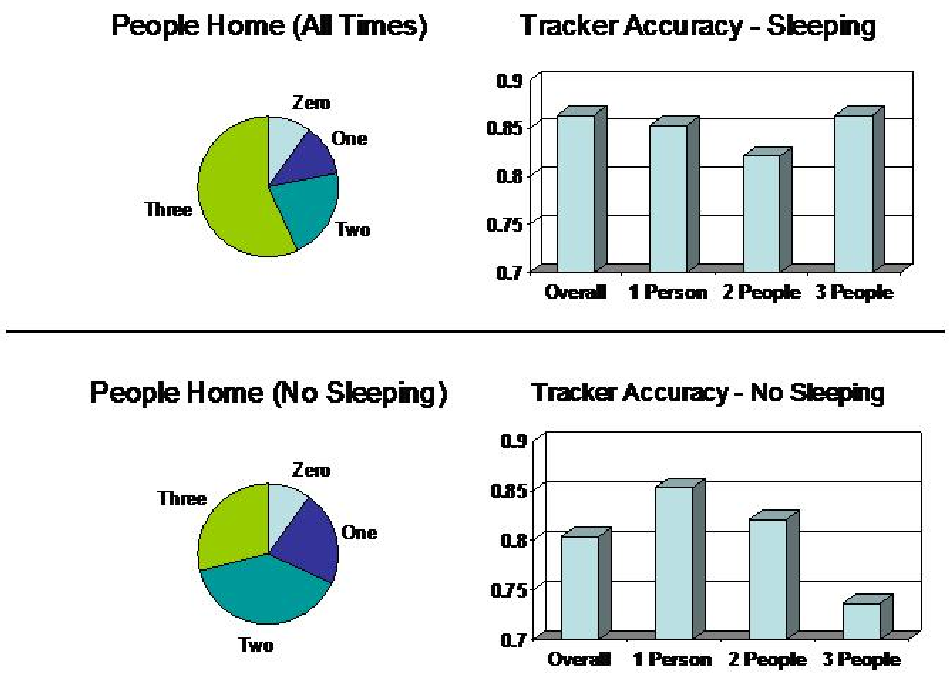

Figure 3.7: Tracking results for STAR experiment # 2.

of the kitchen cabinets and drawers. In these experiments we did not use break beam sensors or pressure mats.

Models. Sensor and motion models were learned before the experiment began (offline) using several days worth of data from when each occupant was home alone.

Measures. Again, location and activity predictions were updated every second and accuracy measured as the number of seconds in which the maximum likelihood predictions of the tracker match the labeled location tag. Results are reported for real-time, online tracker performance.

A researcher hand-labeled the location of each occupant using information gathered by eight wireless keypads. During the experiment, when anyone entered a room with a keypad they pushed the button corresponding to their name. The wireless keypads were placed on the front door, the kitchen, the living room, the study, the downstairs bathroom, the upstairs bathroom, and each of the two bedrooms.

3.6.2 Results

One person experiment. We measured how well a single person was tracked as they moved through the house. The occupant ultimately visited every sensor (including doors, drawers, and the refrigerator) and moved with varying speed and direction. The occupant conducted several common tasks, such as making a sandwich in the kitchen and pausing to use the computer in the study. There were over 1200 sensor readings. Accuracy was 98.2%. We found that even for a single occupant accuracy was never 100% because of occasional lag between entering a room and triggering a sensor.

Two person experiment. In order to understand how the tracker performs with occupants that are co-located versus occupants that are in different places, we scripted two intentionally ambiguous situations in which both occupants shared the same set of anonymous sensors and then separated. The scenario is as follows: two occupants enter the front door thirty seconds apart and move throughout the house without meeting. After fifteen minutes they meet in the living room. One occupant then moves to his bedroom and then returns to the living room. Next, the other occupant leaves to visit his own bedroom and then returns to the living room.

The tracker was accurate for over 98% of the thirty minute experiment. The bulk of the experiment was spent with occupants either moving separately (the first fifteen minutes), or co-located (meeting in the living room). We found near-perfect accuracy when 1) occupants were not co-located and had not recently shared the same sensors, and 2) when occupants were co-located. For example, it is easy to track two people while they watch television together. The difficulty arises when one or both occupants leave the room (the tracker must predict who left). There were two such ambiguous situations in this experiment, and in both cases the ambiguity was resolved as soon as the occupant reached his bedroom. In this case, the motion model contained information about who was more likely to visit a bedroom, and the tracker used it to recover identity. We ran the same experiment using identical generic motion models for both occupants, and found that one of the two recoveries was predicted correctly.

Three person experiment. We measured tracker performance over a five day period for all occupants. There were no guests during this period. When the house was not empty, on average there was one occupant at home 13% of the time, two occupants home 22% of the time, and all three occupants home for 65% of the time. During the experiment every occupant slept in the house. Two of the occupants shared a bedroom and one had a separate bedroom. Every occupant had a separate "study." The tracker used individual motion models for the three occupants. There were approximately 2000 sensor readings each day for a total of 10441 readings. We did not consider the time when no one was home.

On the whole, the tracker correctly classified 84.6% of the experiment. There was no significant difference in accuracy between occupants. The tracker was accurate 85.3% of the time when there was one occupant, 82.1% for two occupants, and 86.4% for three occupants. The system was quite good at tracking sleeping occupants (all three occupants were home each night). Accuracy for three occupants drops to 73.7% when sleeping periods (all data between midnight and 8 AM) are removed.

Figure 3.8: Physical layout of the PlaceLab instrumented apartment.

3.7 Experiment # 3: Instrumented Apartment

In this section, we describe the methodology and results of experiment # 3, in which we performed simultaneous room-level tracking and recognition of several activities of daily living (ADLs) for a single occupant in a home environment. The purpose of this experiment was to determine to what extent our system could recognize several activities of daily living, using only simple sensors.

3.7.1 Study Methodology

The data set used in this experiment was recorded on March 4, 2005 from 9AM to 12PM (four hours total) by researchers at the Massachusetts Institute of Technology working on the PlaceLab project [56]. PlaceLab is described as a "unique live-in laboratory in Cambridge, MA." The data set is called the "PlaceLab Intensive Activity Test Data Set."[3]

Participants. The participant was one female volunteer who was a member of the PlaceLab research team, but not a creator of the core technical infrastructure.

Instrumented Environment. The PlaceLab is an instrumented apartment with eight rooms, including: hallway, kitchen, bedroom, office, bathroom, living room, dining room, and a "powder room" (i.e., a half-bath). See Figure 3.8 for the physical layout of the environment.

[3]The data set is available for download at http://architecture.mit.edu/house_n/data/PlaceLab.htm

Although PlaceLab contains many types of sensors, here we choose to introduce only those sensors used during this experiment, including: motion detectors, contact switches, wireless object movement sensors, water flow sensors, and current sensors. Specifically, there were 18 cameras placed throughout the environment, each of which reported only the amount of motion detected.[4] There were 30 contact switches capable of detecting on/off states placed on doors, knobs, and switches throughout the apartment. Wireless vibration sensors were attached to 106 objects in the environment, and sent a signal whenever that object was moved. There were 14 water flow sensors placed on hot and cold faucets and toilets. 37 sensors measured current flow used by electrical circuits throughout the apartment. See [56] for a more detailed description of PlaceLab hardware.

Procedure. In this experiment the participant was asked to perform a set of common household activities, such as preparing two recipes, cleaning the kitchen, making the bed, and light cleaning around the apartment. The participant was not instructed on the sequence, pace, or concurrency of these activities, and was not limited to performing only these activities. In addition, the participant talked on the phone, answered email, and searched for items – sometimes concurrently. The participant was alone for the entire experiment.

Measures. Location and activity predictions were updated every second using the Bayes filter described previously. In this experiment we used smoothing, or the backward pass, of the Bayes filter. Accuracy was measured as the number of seconds in which the maximum likelihood prediction matched the label. Specifically, we measured accuracy for room-level tracking and eight activities of daily living, including: cleaning a surface, drinking, eating a meal, making the bed, cooking, toileting, using a computer, and using the telephone. These ADLs fall into five important categories for in-home health assessment: cooking, eating/drinking, toileting, socializing, and housework. The activity and location information inferred during this experiment address six of the "top ten" most important ADLs described in Table 2.3 on page 41.

[4]We consider filtered motion data from these cameras to be synonymous with motion detector data.

Activity Name	Accuracy
Idle	79.93 ± 1.26
Cleaning a surface	55.16 ± 4.57
Drinking	89.41 ± 1.79
Eating a meal	7.35 ± 1.55
Making the bed	91.52 ± 19.5
Cooking	81.88 ± 1.46
Toileting	87.80 ± 18.74
Using a computer	65.01 ± 13.37
Using the telephone	63.92 ± 3.69

Table 3.3: Activity recognition accuracy for eight ADLs.

There were four hours of labeled data. We divided the data set into eight thirty-minute segments and used leave-one-out training and testing. Therefore, we report mean accuracy and standard deviation over eight trials, in which we train on 3.5 hours of data and test on the remaining thirty minutes worth. We report accuracy for room-level location estimation and for each of the eight activities of daily living (plus one "none of the above" activity).

3.7.2 Results

Simultaneous tracking and activity recognition. Mean location accuracy over the experiment was $73.30\% \pm 8.9$. Accuracy over the eight ADLs varied widely. The system performed best on making the bed (92%), drinking (89%), toileting (88%), and cooking (82%). Performance fell for using a computer (65%), using the telephone (64%) and cleaning a surface (55%). Accuracy was nearly zero for eating a meal, just 7%. Over the entire experiment, the system correctly recognized when the occupant was "Idle," i.e., performing "none of the above" activities, approximately 80% of the time. See Table 3.3 for all results.

Location accuracy. Location accuracy was lower than we expected for tracking a single occupant in a small environment with hundreds of sensors. We attribute the result to the physical layout of the environment and sensor choice. The open layout of the apartment made tracking a challenge. The dining room and living room were connected openly and the kitchen was separated from the dining room by an island. In addition, many cameras were placed to maximize view and were aimed into multiple rooms. None of the other sensor types were wholly location-specific. For example, water flow sensors and current sensors continue to trigger long after an occupant has walked away from a flushing toilet or a humming microwave. Meanwhile, wireless object movement sensors were often attached to mobile objects with no specific room attachment, e.g., the cordless phone. Location accuracy may improve with different sensor placement or by preprocessing sensor information. For example, by using only derivative information from water flow sensors and current sensors, i.e. reporting only major changes, we could target readings that occur when an occupant manipulates a room-specific appliance or faucet.

Direct sensor associations. Several activities had high accuracy due to well-placed sensors, i.e., direct sensor connections. For example, a wireless moving object sensor was attached to the drinking cup. Another sensor detected when the toilet flushed. On the other hand, eating was very difficult to detect because there were no sensors directly associated with the act. Cleaning also lacked direct sensor associations. Applying simple sensors directly to recognizing specific activities is key to making straightforward probabilistic inference successful.

Room specific activities. We found that accuracy was high for activities which occurred in specific locations. Making the bed, cooking, and toileting always occurred in the bedroom, kitchen, or bathroom, respectively. Cleaning, eating, and using the telephone were not room specific. Using a computer and using the telephone were often performed *concurrently.*

Concurrent activities. In this experiment, using the telephone and using the computer were often performed together.[5] Our system is unable to recognize more than one activity at a time. The system was therefore forced to choose one or the other while both were occurring, which hurt accuracy for both. Any future work in this area should plan to incorporate strategies for dealing with concurrent activities.

Impact of location on activity recognition accuracy. We ran another experiment in which we removed all information about location available to activity

[5]Concurrent activities accounted for 7.5% of the overall dataset.

Activity Name	Configuration # 1	Configuration # 2
Idle	31.04 ± 1.11	68.39 ± 1.20
Cleaning a surface	53.15 ± 4.82	56.25 ± 4.73
Drinking	37.48 ± 14.65	39.13 ± 11.1
Eating a meal	1.21 ± 0.16	7.09 ± 1.35
Making the bed	94.28 ± 12.09	91.52 ± 19.5
Cooking	55.86 ± 1.51	88.41 ± 1.22
Toileting	87.80 ± 18.74	87.80 ± 18.74
Using a computer	38.52 ± 7.92	57.76 ± 12.07
Using the telephone	4.73 ± 1.54	68.61 ± 2.60

Table 3.4: Activity recognition accuracy for two sensor configurations.

recognition. This was accomplished by intializing location uniformly and then skipping the location update step. In other words, at all times during this experiment the system believed that the occupant was equally likely to be in any room.

The results indicate that knowledge of location is key to activity recognition. For the eight ADLs, accuracy drops only an average of 3% relative to accuracy for a system that uses location information. However, the most important difference is in accuracy for the "Idle" activity. This activity alone accounts for 65% of the overall experiment. Here, we see a drop of almost 20% accuracy (from 79.93% to 60.85%) corresponding to a slew of new false positives. We believe that knowledge of location bounds the number of possible activities at any given time, which leads to more conservative estimates of what is happening. Without location knowledge, the system is more likely to predict many disparate activities, leading to more false positives.

Impact of sensor choice on accuracy. We modified the number of sensors available to our system to determine the impact of different sensor combinations. We first ranked which groups of sensors were most important by running the experiment separately using only information from each set of sensors and comparing total accuracy. We found that MITes were most important, followed by motion detectors, contact switches, current sensors, and finally, water flow sensors contributed the least. Next, we ran two experiments, accumulating more sensors in each experiment. First,

Activity Name	Accuracy
Idle	86.25 ± 1.05
Cleaning a surface	49.79 ± 6.57
Drinking	89.41 ± 1.79
Eating a meal	6.13 ± 1.54
Making the bed	91.52 ± 19.50
Cooking	81.17 ± 1.61
Toileting	87.80 ± 18.74
Using a computer	64.26 ± 13.70
Using the telephone	62.54 ± 3.78

Table 3.5: Activity recognition accuracy with a length threshold of ≥ 30 seconds.

we used information from MITes and motion detectors, then we added switches and current sensors. The results are summarized in Table 3.4.

Not suprisingly, accuracy rose as more sensors were added to the system. We found that for the eight ADLs the bulk of accuracy came with just the first set of sensors. Adding new sensors increased accuracy slightly, and for "Making the bed" and "Using the telephone" decreased accuracy slightly. Adding new sensors was mostly useful for straining out noise in the form of false positives. For example, we can see that adding new sensors drastically improves accuracy for the "Idle" state. Thus, we found adding new sensors to be most useful for informing the system of what is *not* happening.

Thresholding to remove false positives. A false positive occurs when the system erroneously predicts that an occupant is engaged in an activity. The clearest indicator that false positives are occurring is low accuracy for the "Idle" activity. In practical use of this system, we believe that false positives are more detrimental than false negatives – the system must gain the trust of the user. Therefore, if the system is going to report information, it should report the right information – even if some activities are missed.

To reduce false positives, we implemented a simple thresholding approach which removed all predicted activities that lasted less than 30 seconds. Such activities were changed to the "Idle" activity. The benefit is that the rate of false positives should drop, as most activities last longer than 30 seconds. The drawback is that accuracy for short duration events should also be reduced. As we can see in Table 3.5, accuracy for most activities remains the same, but the accuracy for the "Idle" state rises significantly, corresponding with fewer false positives. Accuracy for "cleaning a surface" drops slightly, because this is a short duration activity. Overall, a quick thresholding over activity duration is able to significantly reduce the number of false positives detected by the system.

HMMs Vs. HSMMs. Many researchers have explored using more complex HMM models for activity recognition. Such models include layered HMMs [87] and more recently, hidden semi-Markov models (HSMMs) [41]. We also implemented an HSMM approach, but we did not see a significant gain in accuracy. HSMMs explicitly model state durations, however, our approach uses a single state for each activity. Without modeling specific steps of an activity, we believe the benefits of an HSMM are wasted. However, due to the popularity of semi-Markov models for activity recognition, we did choose to extend our activity rating algorithm in chapter 5 to handle HSMMs.

3.8 Discussion

We have shown that tracking multiple occupants in a home environment and recognizing activities of daily living is feasible via a set of simple sensors. In summary:

- We found that highly predictive motion models improve accuracy, regardless of whether occupants behave similarly. In practice, the differences between motion models show up in private areas, like bedrooms and bathrooms, or during personal activities, like sitting in a favorite easy chair. The bigger these differences, the easier data association becomes and the more accuracy improves.

- Parameter learning is straightforward when an occupant is alone, however, occupants behave differently in groups. Learning models online could mitigate this discrepancy. In simulations, we found that the accuracy of models trained online falls as the number of occupants rises. One promising solution would be to combine online and offline approaches.

- The number of particles required depends on the complexity of the data association problem. More particles are necessary for environments with many occupants and sensors. In particular, more particles are necessary for situations in which multiple co-located occupants separate. We found negligible accuracy improvements after twenty or so particles, even for up to five occupants. This number may change depending on the efficiency of the particle filter approach and the data association proposal scheme.

- More sensors will increase accuracy, regardless of the number of occupants. A low sensor density contributes to significant periods of time between readings (especially with only one occupant). During these "quiet" times no new information arrives to help the tracker recover from mistakes (such as the lag between entering a new room and triggering a sensor). In experiment # 1, we found that motion detectors were the most active sensors, and a lack of them hurts tracking accuracy the most. In experiment # 3, we found that wireless object movement sensors (particularly hundreds of them) greatly improve activity recognition accuracy.

- More occupants will decrease accuracy, particularly if parameter learning is performed completely online and motion models are generic. The accuracy suffers most when data association becomes difficult, i.e., immediately after co-located occupants separate. In general, accuracy is high for co-located occupants and for occupants who have not come into contact with other for some time.

- Knowledge of room-level location improves activity recognition accuracy over several activities of daily living. In addition, knowledge of location greatly reduced false positives, i.e., erroneously reported activities, by limiting the number of likely activities to those common to a room.

3.9 Conclusion

In this chapter, we reviewed a variety of related work in people tracking and activity recognition. We outlined our sensor configuration and described why we chose to use simple sensors. To provide information crucial to automatic health monitoring, we introduced the STAR problem and showed how information from anonymous, binary sensors may be used to provide simultaneous location estimation and activity recognition for multiple occupants in a home environment. We evaluated our approach on three data sets, collected from simulation and two instrumented environments.

Chapter 4

Data Collection in the Home

Practical in-home health monitoring technology depends upon accurate activity inference algorithms, which in turn often rely upon labeled examples of activity for training. In this chapter, we describe a novel, unsupervised technique in which contextual information gathered by ubiquitous sensors is used to help users label a multitude of anonymous activity episodes.[1] This technique, called the context-aware recognition survey (CARS), appears as a game-like computer program in which users attempt to correctly guess which activity is happening after seeing a series of symbolic images that represent sensor values generated during the activity. Our approach is valuable because it is practical: Users may label each others' data at any time without additional instrumentation or any interruption to daily routine via an easy-to-use, video game-like interface.

4.1 Introduction

Many ubiquitous computing applications depend on knowledge of how people behave in their environments. Fortunately, pervasive computing applications implicitly gather a valuable *context history* as they collect and store sensor data over time; unfortunately, this vast amount of data is of limited use without labels that explain what was happening in some way. Although potentially costly to collect, labeled examples

[1] This chapter is a revised version of [119]

of activity can both improve design decisions and help machine learning techniques to recognize and predict activity. The work described in this chapter is motivated by a need to acquire labeled examples of activity in the home without burdening occupants or disrupting their daily routines.

We describe the *context-aware* recognition survey (CARS), which helps users to label anonymous activity episodes by displaying contextual information collected by ubiquitous sensors such as contact switches, pressure mats, motion detectors, and RFID readers. Drawing on recent research in practical home monitoring systems, game-based image-labeling techniques, and data labeling techniques [116, 120, 5, 16], we designed a game-like multiple choice test that displays low-level sensor readings as colorful symbols and descriptive text. Users answer the questions with the goal of correctly labeling the activity being depicted. This approach allows anyone to label the data at any time, without requiring additional hardware (beyond the original sensor infrastructure) or causing additional interruption to daily routine. We believe these properties are necessary if a practical in-home data labeling system is to be used on a wide scale by non-experts.

We report two experiments: One in which users (N=10) performed a subset of tasks in an instrumented environment and completed a CARS approximately one week later, and another in which users (N=20) completed a more complicated, real-world version of a CARS, providing labeled data for over 25 activities performed by complete strangers. Results from the first experiment draw upon statistical analyses of test performance and a post-test questionnaire. We present a comparative analysis of how well users performed on their own activities (which they may remember) vs. the activities of others (which they have never seen) vs. counterfeit examples of activity. In the second experiment we explore whether a completely unsupervised system is feasible. In this study, episodes of activities were automatically segmented from the entire stream of data and chosen for labeling according to algorithms borrowed from the active learning literature. We summarize our results with a series of "lessons learned," and suggestions for design decisions for any future CARS. These findings have value that extends beyond the specifics of our own system, both in terms of the design of data collection systems and research on the abilities and needs of human labelers.

In the this section we introduced the data labeling problem and outlined our solution. Next, we describe related work in the area of data labeling. We then describe our implementation of a context-aware recognition survey and demonstrate results over two experiments in which participants used our system to label data generated in two different instrumented environments. Finally, we provide a critique of our approach and make suggestions to future researchers who may wish to build on our technique.

4.2 Related Work

Several standard classes of methods exist for collecting data about daily activities, including one-on-one or group interviews, direct observation, self report recall surveys, time diaries, and the experience sampling method (ESM) [13, 57]. While direct observation and annotation of video data are often reliable, they are prohibitively time-consuming and insensitive to privacy concerns. In interviews and recall surveys, users often have trouble remembering activities and may censor what they do report. Time diaries may reduce recall and selective reporting bias, but require a commitment from the user to carry around (and use) the diary. Experience sampling techniques use a prompting mechanism (e.g., a beep) to periodically ask the user for a self-report. Prompts may interrupt activities and must be carefully delivered in order to avoid annoying the user [57]. All of these methods require the participation of the person who performed the activity and others may require outside help as well (e.g., interviewers or annotators). In contrast, our approach is designed to never interfere with the occupant and to be performable by people who may have never experienced or witnessed the activity they are labeling.

Existing data collection schemes increasingly exploit an underlying sensor infrastructure. The electronic experience sampling method (ESM) collects data by using a portable computing device (e.g., a PDA or cell phone) to prompt users and to collect their responses. Because prompting can interrupt activities and quickly become an annoyance [13], *context-aware* ESM approaches have been developed which use additional sensor information to choose the best time to prompt for data [100]. The benefit is that prompts should only occur when they are perceived to be least invasive and when the information to be gained is maximally beneficial. The context-aware recognition survey has great potential to mesh seamlessly with established context-aware ESM techniques. There are some immediate advantages: 1) The CARS could be used to "fill in the blanks" from missed or ignored ESM prompts, 2) labels collected via ESM could be verified or supplemented by CARS at a later date, and 3) ESM data could be used to improve automatic episode recovery in the CARS system.

The data collection technique that is most similar to the CARS is called "image-based experience sampling," which mitigates forgetfulness common to recall surveys by using cues such as photo snapshots [57]. In previous work, researchers have used snippets of video, audio, and photo snapshots to help remind users who are performing a recall survey of what activities they performed. This approach is desirable because it reduces recall bias, avoids interrupting daily routines, and does not require the addition of wearable (or carryable) sensors. (Ideally, the cues are chosen automatically, i.e., without human intervention.) Our approach is desirable for the same reasons, however, we do not use information from cameras or microphones. Instead, our approach maintains privacy by converting low-level sensor values into non-private, generic symbols. The added benefit is that these episodes can be labeled by anyone at any time.

It has been shown that the human desire to be entertained can be tapped for labeling data. Researchers at CMU recently demonstrated the potential of using "human cycles" to label images on the Internet [5]. In the "ESP Game," two online players examine the same image and independently type descriptive words. The result is a double-blind test that rewards players who come up with the same answers. Our method currently engages a single user off-line, however, the game-like setup lends itself well to the transition to an online video game. We do not pursue this approach in these experiments, however, we forsee advantages in the form of more reliable labels which have been annotated many times by different users (which may be especially important when it is infeasible or inconvenient to ask the performer of the activity to label).

Finally, existing sensor infrastructure is increasingly prevalent in the home. The cost of sensors and computation has dropped to the point that consumer-ready sensor installations are becoming possible. A study conducted at Intel Research Seattle found that participants were able to install faux sensors according to a simple instruction sheet, without expert supervision [16]. Meanwhile, other researchers are designing and deploying real sensor installations for the home [14, 120]. Simple, do-it-yourself sensors are likely to constitute the first examples of actual in-home monitoring systems. These sensors are chosen specifically for real homes; they are low-cost, affordable, and not perceived as invasive. We have taken the following as a design challenge: to use output from an existing sensor infrastructure to help users label activities. We envision the CARS as a tool usable by non-experts to "train" their in-home monitoring systems.

4.3 Approach

The key idea of the context-aware recognition survey is to use contextual information collected by ubiquitous sensors to provide an augmented recall survey that can be performed by anyone at any time, regardless of who performed the activity or how the sensors were configured. The technique consists of the following steps: 1) Sensor readings are collected over time and stored, 2) sensor readings are automatically segmented by activity into episodes (called *episode recovery*), 3) episodes are converted into a series of generic, highly descriptive images, and 4) episodes are labeled by users in a game-like computer-based recognition survey. Afterwards, the labeled episodes may be used to train machine learning algorithms or to improve design decisions for pervasive computing applications.

We describe two experiments; in both experiments a series of unlabeled episodes were translated into symbolic images and short lengths of text via a tool called the Narrator [116]. The Narrator is a finite state machine that parses low-level sensor information and generates a concise, readable summary. For this study, a simplified version of the Narrator mapped each change in sensor values to an image and caption.

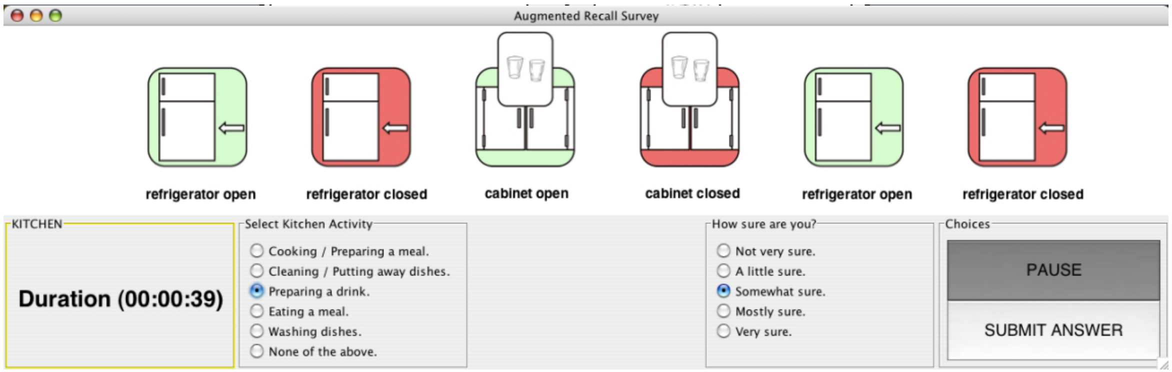

Figure 4.1: Screenshot of CARS for experiment # 1.

The source of the sensor reading (e.g., contact switch, motion detector, RFID tag) is inconsequential, so long as it can be mapped to a generic symbol. Each new sensor configuration must be hand-mapped to a standard set of generic images that may be shared by many other instrumented homes. The resulting series of images (and captions) describe an activity episode in simple, symbolic terms. These episodes are then presented to users in the form of a recognition survey.

4.4 Experiment # 1

In this experiment we gauged the feasibility of using a context-aware recognition survey for labeling data in the home by asking ten participants to generate data in an instrumented home environment and then having them each complete a CARS approximately one week later. See Figure 4.1 for an example screenshot of the program used in this experiment.

4.4.1 Study Methodology

Participants. Participants in this study were 10 adult volunteers who were recruited from the university and from the community. Participants ranged in age from 25 to 32 years, and the sample was 50% female and 50% male. Participant background varied, ranging from librarians to engineers.

Instrumented environment. This study occurred in the lead author's home. A kitchen and bathroom were instrumented with two types of anonymous, binary sensors: magnetic contact switches and pressure mats. In the kitchen, contact switches were installed on the refrigerator, freezer, and microwave doors, as well as on the trash can, liquid soap dispenser, stove top burner, hot and cold water faucet knobs, two cupboards, and two drawers. A pressure mat was placed under the chair at the kitchen table. The refrigerator, freezer, cupboards and drawers used in the experiment were clearly labeled with their contents (e.g., ice, plates, cups). In the bathroom, magnetic contact switches were installed on the liquid soap dispenser and hot and cold water faucet knobs, and a pressure mat was placed on the floor in front of the sink. All sensors interfaced with a desktop computer via an extended parallel port. Sensors were polled every second and values were stored in a mySQL database.

Activity recording. Participants participated one at a time. First, participants were asked to perform several activities in the instrumented kitchen and bathroom. They were informed that they would be asked to recall their activities in a "quiz" later that week. The locations of objects needed to perform the activities were clearly labeled and each participant was given an initial tour of the instrumented rooms. Participants were instructed to choose and perform a subset of several kitchen tasks (which were also posted on the refrigerator as a reminder). The kitchen tasks were: Prepare a cold drink, prepare either a sandwich, a fried egg, or a microwave pizza, eat the meal, wash dishes and put them away, and throw away any trash. During the bathroom portion, participants were given a toothbrush and were instructed to brush their teeth and then perform two of three tasks: Wash their face, wash their hands, and comb their hair. An observer time-stamped the start and end points of each activity using a laptop computer. Participants were instructed not to speak with other participants about which tasks they had performed.

Context Aware Recognition Survey. We presented our computer-based recognition survey as a "game" in which the goal was to correctly guess which activities were happening given only the sensor readings collected from the kitchen and bathroom environments. The contextual information gathered by the sensors was hand-segmented using the start and end points time-stamped by the observer, and converted into

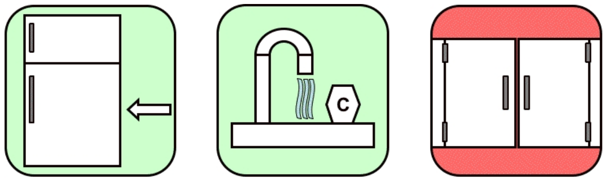

Figure 4.2: Symbols: (a) Refrigerator open, (b) water on, (c) cabinet closed.

episodes of images and text via the Narrator program [116]. We attempted to address the implications of improperly segmented (i.e., malformed) episodes by asking users to label *counterfeit episodes*. Counterfeit episodes were hand-made by intentionally choosing incorrect activity durations and by introducing noisy sensor readings (i.e., dropped readings and spurious readings). Several counterfeit episodes (which did not correspond to any activity) were generated by hand. The rest of the episodes were "real" in that they were generated by participants performing activities in the instrumented environment.

Each episode consisted of a series of scrolling images that had red or green backgrounds, depending on whether that object was turned on or off. See Figure 4.2 for example symbols, or Appendix F on page 151 for the complete list. The word "kitchen" or "bathroom" was presented with each episode to indicate the location where the activity took place. The total duration of the episode was also displayed (no other timing information was included). Participants were able to pause the scrolling pictures, but were not able to replay an episode. After viewing an episode, participants were asked to select from a multiple choice list of every possible kitchen or bathroom activity (depending on which room the activity occurred in) plus a None of the Above answer. Participants were also asked to rate how confident they were about their choice on a scale of one to five.

Participants were administered the CARS on a laptop computer a mean of 5 days following the activity recording (Range=2-7 days, SD=1.63 days). Each participant was presented with two sets of 12 activity episodes, which we call the "self" set and the "other" set. The self set contained 8 episodes from the participant's own activities and 4 counterfeit episodes. The other set contained 8 episodes of someone else's activities and 4 counterfeit episodes. Participants were informed of which sets were self or other. The survey administration was counterbalanced, with half of the participants presented the self set first, and the other half with the other set first.

Subjective Experience Questionnaire. Immediately after completing the computer-based CARS, participants completed a brief paper and pencil questionnaire about their subjective experience.

4.4.2 Results

1. Participants successfully identified 82% of the 24 total episodes (M=19.60, SD=3.47). Assessing confidence in their selection on the Likert scale of 1-5 (1=Not Very Sure and 5=Very Sure), participants reported being Mostly Sure (M=3.96, SD=1.03) across all of the episodes. Overall, **participants were able to successfully label most activities with confidence.**

2. The number of days between activity performance and activity recall ranged from 2 to 7 (M=5.00, SD=1.63) and was not significantly correlated with total performance scores, r=.27, p=.44. Therefore, we found that **the amount of time between activity and recall did not significantly affect performance in our sample.**

3. Ignoring counterfeit episodes, performance on the self section (M=7.10, SD=1.29) and the other section (M=7.10, SD=.99) was identical, with participants correctly identifying 89% of the 8 possible episodes. There was also no significant difference in participants' confidence ratings in their identification of their own activities vs. someone else's. Overall, **participants were equally good at labeling their own or other peoples' real activities.**

4. Although there were no significant differences in overall performance on self vs. other episode identification, differences did emerge for the most difficult to identify episodes. Due to limited sensor granularity, several bathroom tasks (e.g., face washing vs. hand washing) were only recognizable almost solely from memory. The number of errors on these tasks were low in the self section, but more than doubled in the other section. This indicates that **the importance of memory increases for**

hard-to-recognize activities.

5. Participants had better performance on the 16 real episodes (M=14.20, SD=1.47) than on the 8 counterfeit episodes (M=5.40, SD=2.32). This proportional difference (67% for fake, 89% for real) was significant, $t(9)$=-2.80, p<.05. The mean confidence rating was also slightly higher for real (M=3.02, SD=.37) vs. counterfeit episodes (M=2.82, SD=.43), although this difference was not statistically significant, $t(9)$=1.43, p<.19. This indicates that although participants may have been unaware of this deficit, **counterfeit episodes were often mislabeled as real episodes.**

6. The order of test administration (self then other, or vice versa) did not impact overall performance or performance on self vs. other sets, but did impact performance on the identification of counterfeits. Participants who completed the self section first were significantly better at detecting fake episodes in the other section, $t(8)$=2.36, p<.05. We hypothesize that users become better at spotting counterfeit episodes as they gain experience. In other words, **practice makes perfect.**

7. Participants who completed the other section first did not perform significantly better at spotting counterfeits on the self section. This could be because experience gained during the other set did not overcome the inherent gains offered by remembering which activities did or did not occur. We believe that this may be because **remembering activities makes it easier to spot counterfeit episodes.** One participant writes, "*I could remember the steps I took...and I knew it was me.*"

8. Performance on each episode was significantly related to participants' rating of their own confidence level in their selection for that episode, r=.16, p<.05. There was also a significant difference between mean confidence level on correct (M=3.03, SD=1.03) vs. incorrect (M=2.61, SD=1.06) selections, $t(238)$=2.39, p<.01. This suggests that **user confidence ratings are potentially a useful measure of whether the episode was correctly labeled.**

9. There was no difference found between confidence ratings on self vs. other activity identification, however, in the follow-up survey, participants reported that in their overall experience of the test it was more difficult to identify someone else's activities than their own $t(9)$=-1.95, p<.10 (significant at the trend level). It appears that **participantively, participants feel it is easier to label their own**

activities.

10. Participants reported that their participantive experience of the test was positive. Participants reported that the symbolic images were “pretty easy” to “very easy” to understand on the Likert scale of 1-5 (M=4.70, SD=.48). Open-ended questions that asked what participants liked and did not like about the CARS also indicated that participants had a positive experience, with participants reporting that they liked the color-coding for off and on, and that the images were “cute,” “clear,” or “easy to understand.” All in all, **participants enjoyed taking the CARS.**

11. Total performance scores ranged from 11 to 23 correct identifications (out of 24 possible), with 90% of the participants correctly identifying 18 or more of the episodes. Most participants were fairly adept at labeling, while a few performed significantly worse. Obviously, not everyone makes a good choice for a labeler. However, in the follow-up survey several high-performing participants requested the ability to speed up the scrolling. This indicates that when designing the CARS **researchers should plan for the varying abilities of different participants.**

4.5 Practical Considerations

We identified three main weaknesses in our initial CARS implementation: 1) We used low-granularity sensors (e.g., contact switches) which made some activities impossible to recognize, except from memory; 2) we depended on a human to hand-segment the data into episodes; and 3) we had no mechanism for optimally choosing the order in which episodes should be labeled, and thus minimizing the number of necessary questions. In this section, we briefly describe our solutions in these areas before going into more detail in the next section.

4.5.1 Higher Granularity Sensors

In experiment # 1, we found that our choice of simple sensors did not provide sufficient granularity for users to confidently label certain activities. For example, it was particularly difficult to tell the difference between washing hands and face (in

Figure 4.3: Pictures of (a) The iBracelet, a wearable RFID reader, (b) tagged objects.

this case, the only information was that a pressure mat in front of the sink was stepped on and cold water was used). In response, we integrated higher granularity RFID sensors, specifically the iBracelet [92], developed at Intel Research Seattle. The iBracelet reports a unique ID number for every tagged object that is touched in the environment. See Figure 4.3 for a picture of the iBracelet.

4.5.2 Automatic Episode Recovery

An attractive aspect of the context-aware recognition survey is the fact that it is completely unsupervised (aside from the user labeling step). In our previous study, however, we hand-segmented the stream of sensor readings generated by the user. To automate this step, we performed unsupervised episode recovery using HMM models bootstrapped with common-sense information mined from the Internet (an approach pioneered at Intel Research Seattle [92]). The key idea was to train rough HMM models with information "scraped" from instructional web pages, and then to use these models to identify the segments between activity episodes.

4.5.3 Active Learning for Episode Selection

In most situations, the amount of unlabeled data vastly outweighs the amount of labeled data. Even after automatic segmentation into unlabeled episodes, much of this data may be useless. For example, an unlabeled episode could have been segmented incorrectly, riddled with noisy sensor readings, or it may describe an activity that

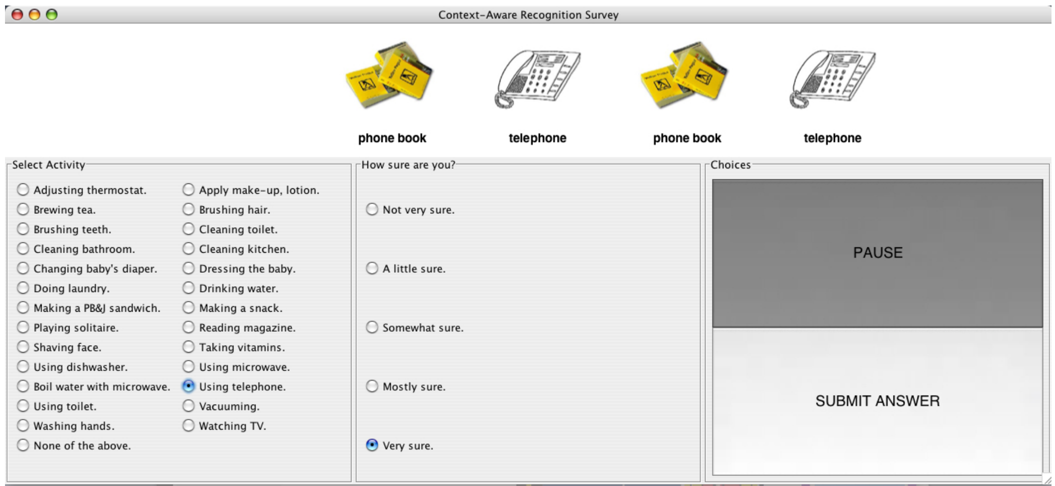

Figure 4.4: Screenshot of CARS for experiment # 2.

already has a well-trained model. In addition, the amount of time it takes a human to label an episode represents a scarce resource – there are often far too many unlabeled episodes to ask a human to label all of them. Clearly, there is a need to intelligently select the best episodes for labeling, in order to 1) minimize the amount of labeling necessary and 2) to maximize the usefulness of labeled examples for training activity models. We borrow algorithms from the field of *active learning*, specifically the *query-by-committee* approach, to provide a solution.

4.6 Experiment # 2

In this section, we describe an experiment in which we ask participants to complete a refined version of the context-aware recognition survey that incorporates data from higher granularity sensors, automatic episode recovery, and active learning for episode selection. We report results from twenty participants who completed this CARS. See Figure 4.4 for an example screenshot of the program used in this experiment.

4.6.1 Study Methodology

Participants. Participants in this study were 20 adult volunteers recruited from the university and from the community to complete a computerized context-aware recognition survey. Participants ranged in age from 22 to 45 years, and the sample was 45% female and 65% male. Participant background varied, ranging from airline pilots to architects.

Instrumented environment. Participants were asked to complete a CARS that used data generated by complete strangers in a previous study [92]. None of the participants participating in this experiment were involved in generating this data. In the previous study, a real home in the Seattle area was selected and instrumented with over 100 RFID tags. Objects as diverse as faucets and remote controls were tagged. Figure 4.3 illustrates the RFID infrastructure that was assumed. On the left is a bracelet which has incorporated into it an antenna, battery, RFID reader, and radio. On the right are day-to-day objects with RFID tags (battery-free stickers that currently cost 20-40 cents apiece) attached to them.[2] The bracelet-mounted reader constantly scans for tags within a few inches of the hand. When the wearer of the bracelet handles a tagged object, the tag on the object modulates the signal from the reader to send back a unique 96-bit identifier (ID). The reader can then ship the tag ID wirelessly to a base computer which can map the IDs to object names.

Activity recording. In the previous study, 9 non-researcher participants with a wearable RFID reader performed, in any order of their choice, 14 ADLs each from a possible set of 65; in practice they restricted themselves to 26 activities over a single 20 to 40 minute session. There were no interleaved activities and a written log was used to establish ground truth. See Appendix H on page 155 for the complete list of tagged objects and activities performed.

Automatic Episode Recovery. We used bootstrapped HMMs for automatic episode recovery, however, there are dozens of other successful techniques available to solve the same problem. See [65] for a survey of segmentation techniques and [96] for an introduction to HMMs. We chose to use the HMM approach because it

[2]We assume that participants or their caregivers will tag a multitude of objects in the home environment. In experiments we have tagged over a hundred objects in a real home in a few hours.

is extremely straightforward, fast, and required minimal human intervention. Also, in other research we are already using similar HMM models for activity recognition. Segmentation accuracy could probably be improved with a more high-powered segmentation technique, however, we intentionally choose to keep segmentation simple and rely on active learning to filter out badly segmented episodes. Ideally, the active learning mechanism should never choose to show a badly segmented episode to the end user, as long as there are some unlabeled and well-segmented episodes still available.

A hidden Markov model (HMM) was trained on information gathered from the Internet. This data-mining process used word appearances on "how to" websites to compute the probability that an object was used during each activity. Using the website www.ehow.com, 65 different web pages were chosen that described activities ranging from using the microwave to using the toilet. Each web page contained some number of nouns that matched tagged objects in the environment. For example, we tagged the microwave and the term "microwave" appeared 7 times on the page entitled "How to Boil Water in the Microwave."

From this mined information we assembled an HMM with one state for each of the 65 possible activities, and a set of observations composed of the set of mined objects, pruned to include only those which we knew were in our set of deployed tags. The observation probabilities were then set to normalized values of the mined probabilities. We set the HMM's transition probabilities to reflect an expected number of observations (5) for each activity, as well as a uniform probability of switching to any other activity. See [92] for a detailed description of this data-mining process.

Next, for each of the 9 sensor traces (one for each participant) we used the Viterbi algorithm [96] to compute the most likely sequence of labels for each sensor reading (i.e., each object touched). We then simply segmented the labeled trace into contiguous sequences of the same label. See Table 4.1 for an example of this data.

We used the P_k metric to measure segmentation accuracy [17]. The P_k metric is the probability that two observations at a distance of k from one another are incorrectly segmented. As such, it can be thought of as the error rate for the segmentation, and 1 - P_k can be thought of as the segmentation's accuracy. The value of k is set to one half

RFID tag	HMM label	Segment #
14	9	0
23	9	0
21	9	0
14	9	0
26	5	1
29	5	1
12	7	2
11	7	2
13	7	2

Table 4.1: Example of automatically segmented data.

of the average segment length (in our case $k = 3$). The P_k score for our segmentation using only the mined parameters (i.e., with no training) is 29.7, indicating that we should expect to be able to segment sensor traces in a completely unsupervised manner with higher than 70% accuracy. This is a promising indication that bootstrapped HMM models can be used to perform unsupervised episode recovery.

Active learning for episode selection. After automatic episode recovery, we were left with 145 unlabeled episodes. We decided to present only 30 episodes to each user, resulting in an experiment that usually lasted from 10 to 15 minutes. (Practically speaking, it was infeasible to expect participants to label all 145 episodes.) In order to select the "best" 30 episodes, we borrowed a well-known algorithm from the active learning community, called *query-by-committee* (QBC) [103].

The key idea behind QBC is to 1) sample a "committee" of slightly different HMM models (each constituting a "member" of the committee); 2) use each sampled model to label every episode (each member "votes"); and then 3) choose to label episodes which have high entropy between committee members' labels (the episodes where members most disagree). Finally, we use the newly labeled episode to re-train the original HMM model, before repeating the same process for the next episode. The algorithm is implemented as follows:

Step 1. We wish to sample a committee of models from the posterior parameter distributions $P(\alpha_i = a_i|S)$ of the original HMM model, given HMM statistics S (we

Episode #	Label one	Label two	Label three	Vote entropy
1	8	8	8	0.0
2	2	3	2	.667
3	4	1	5	1.0
4	3	6	1	1.0
5	7	7	4	.667

Table 4.2: Entropy between 3 committee members for 5 episodes.

assume a uniform prior). We note that the parameters of an HMM represent a set of multinomial probability distributions. Let $\{u_i\}$ denote the set of possible values of a given multinomial variable (e.g., the possible observations for a given activity), and let $S = \{n_i\}$ denote a set of statistics extracted from the training set, where n_i is the number of times that the value u_i appears in the training set. We denote the total number of appearances of the multinomial variable as $N = \sum_i n_i$. The parameters whose distributions we wish to estimate are $\alpha_i = P(u_i)$.

The maximum likelihood estimate for each of the multinomial's distribution parameters, α_i, is $\hat{\alpha}_i = \frac{n_i}{N}$. We smooth this estimator to compensate for data sparseness:

$$\hat{\alpha}_i^S = \frac{(1-\lambda)n_i + \lambda}{(1-\lambda)N + \lambda v}, \tag{4.1}$$

where $\lambda << 1$ is a smoothing parameter controlling the amount of smoothing (set to .05 in our experiments), and v is the number of possible values for the given multinomial, i.e., the number of possible observations (in our case 68, for the full list see Table H.2).

The posterior $P(\alpha_i = a_i|S)$ is a Dirichlet distribution [58]. To simplify implementation, we assume that a multinomial is a collection of independent binomials, each of which corresponds to a single value u_i of the multinomial. For each such binomial, we sample from the truncated normal approximation for the *smoothed* estimate, with mean $\mu = \hat{\alpha}_i^S$ and variance $\rho^2 = \frac{\mu(1-\mu)}{N}$. Here, n_i equals the number of times value u_i appears, N equals the total number of appearances of the multinomial variable, and v equals the number of possible values for the multinomial. Afterwards, we renormalize the sampled parameters so that they sum to 1.

Figure 4.5: Symbols from left to right: (a) Faucet, (b) bleach, (c) toothbrush.

Step 2. The next step is to label each episode according to the parameters of our newly sampled models. This is done in the traditional method, via the Viterbi algorithm [96]. Next, we must calculate the entropy between different answers. We denote the number of committee members assigning a label c for input example e by $V(c, e)$. Now, we can measure the disagreement between committee members by the entropy of the distribution of labels "voted for" by the committee members. This measure, called *vote entropy*, quantifies the uniformity of classes assigned to an example by many different committee members. The vote entropy is normalized by a bound on its maximum possible value, $\log(\min(k, |c|))$, resulting in a value between 0 and 1 (corresponding to complete agreement or disagreement, respectively). The formula for *normalized vote entropy* is as follows:

$$D(e) = -\frac{1}{\log \, min(k, |C|)} \sum_{c} \frac{V(c,e)}{k} \log \frac{V(c,e)}{k}. \tag{4.2}$$

See Table 4.2 for an example of this data.

Step 3. We use a simple selection criteria called *randomized selection* to decide which episode to present to the user. By this technique, an example is selected randomly, weighted by the vote entropy (higher vote entropy corresponds to a higher probability of selection).

4.6.2 Context Aware Recognition Survey

Once again, we presented our computer-based recognition survey as a "game" in which the goal was to correctly guess which activities were happening given only the sensor readings collected from the instrumented home.

The automatically segmented episodes were again converted into episodes of images and text via the Narrator program [116]. Each episode consisted of a series of scrolling images that corresponded to which object had been touched.[3] See Figure 4.5 for example symbols, or see Appendix G on page 153 for a complete list of symbols. No location information was given, although location was often obvious when occupants had touched objects such as "bathroom door," or "microwave." The total duration of the episode was not displayed, nor was any other explicit timing information. Participants were able to pause the scrolling pictures, but were not able to replay an episode. After viewing an episode, participants were asked to select from a multiple choice list of every possible activity (of which there were 26) plus a "None of the Above" answer. Participants were also asked to rate how confident they were about their choice on a scale of one to five.

Participants (N=20) were administered the CARS on a laptop computer. All participants were given a "walkthrough" of the program, and a brief description of how to answer ambiguous questions. See Appendix I on page 158 for a copy of the information sheet. A group of 10 participants was presented with 30 activity episodes chosen randomly from 145 possible. Another group of 10 participants was presented with 30 activity episodes chosen according to the query-by-committee algorithm described above. We call these the "random" and "active" question sets, respectively.

4.6.3 Results

1. On average, participants were able to answer about 20 of 30 questions correctly (M=19.4, SD=3.38), although scores ranged from 14 to 26. There was no significant difference between scores on actively chosen questions sets versus randomly chosen data sets. This indicates that **participants were able to successfully label over 25 different activities completed by complete strangers in an instrumented environment.**

2. Participants assessing their confidence in their selections on the Likert scale of 1-5 (1=Not Very Sure, 5=Very Sure) reported being Mostly Sure across all questions

[3] Using Google, we assembled 68 prototype object-symbols in a few hours.

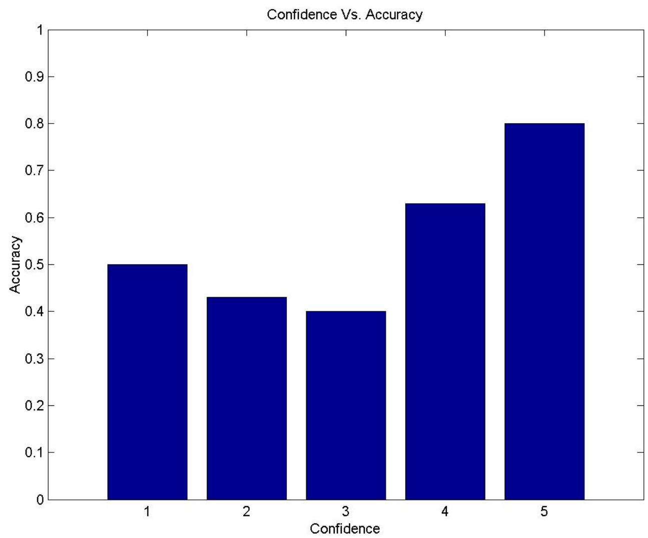

Figure 4.6: Relation between confidence and labeling accuracy.

(M=4.02, SD=1.07). Confidence between active and random question sets was not significantly different. Also, there was no significant change in confidence as the number of questions increased (i.e., participants did not seem to become more or less confident over time). Labels that were associated with a confidence of Not Very Sure were wrong 50% of the time, A Little Sure 57% of the time, Somewhat Sure 60% of the time, Mostly Sure 37% of the time and Very Sure 20% of the time, on average. **This corroborates earlier findings indicating that participants' confidence ratings can be used to predict labeling accuracy**. See Figure 4.6.

3. Note that the 145 possible episodes available for labeling were segmented automatically. As a result, 17 of these episodes were so badly segmented that they did not correspond to any activity (i.e., the correct label for these episodes is None of the Above). During the study, participants were instructed to choose None of the Above for episodes which seemed incomprehensible or which seemed to represent more than one activity at the same time.

On average, participants chose None of the Above about 4 times out of 30 questions (M=4.4, SD=3.5), with a range from 0 to 14. Participants were significantly more likely to choose None of the Above on the randomly chosen question set (M=6.2, SD=3.91) than on the actively chosen question set (M=2.6, SD=1.84), with a factor of $t(20)$=2.63, p=.017. There was also a difference between the number of "bad" episodes (i.e., episodes that corresponded with no activity) presented by the active and random question sets. In the active question set there were significantly fewer of these questions (M=2.5, SD=1.27) than in the random question set (M=3.7, SD=1.34), with a trend level significance factor of $t(20)$=2.01, p=.054. This indicates that **the active learning algorithm intentionally avoided presenting participants with badly segmented episodes**.

4. We recorded the number of seconds that each participant spent before choosing an answer for each question. The timer started as soon as the entire question had been displayed and ended when the participant pressed the "Submit" button. This number was recorded as "0" when participants answered before the entire question had been shown (which they were instructed not to do).

We found that for the actively chosen questions there was a fairly even amount of time spent for every question. However, for the randomly chosen questions participants spent considerably longer on each of the first three questions (M=10, SD=12) than on each of the last 27 questions (M=5.1, SD=5.5). This difference was extremely statistically significant, $t(300)$=3.96, p<.0001. We hypothesize that this data represents a steep "learning curve" which may have been exacerbated by the fact that episodes were chosen randomly, with participants more likely to be confronted initially with badly segmented questions which required more time to figure out. In addition, we found a significant difference between overall amount of time spent on actively chosen questions (M=3.7, SD=4) vs. randomly chosen questions (M=5.6, SD=6.7), with participants recorded spending an average of 2 seconds less on the actively chosen questions, $t(600)$=4.22, p<.0001. This indicates that **episodes in the active question set were labeled faster, possibly because they were easier for participants to answer**.

5. Time spent answering a question was significantly correlated to whether the question was answered correctly, with incorrectly answered questions usually taking approximately 2 seconds longer to answer on average, with $t(600)$=20.85, p<.0001.

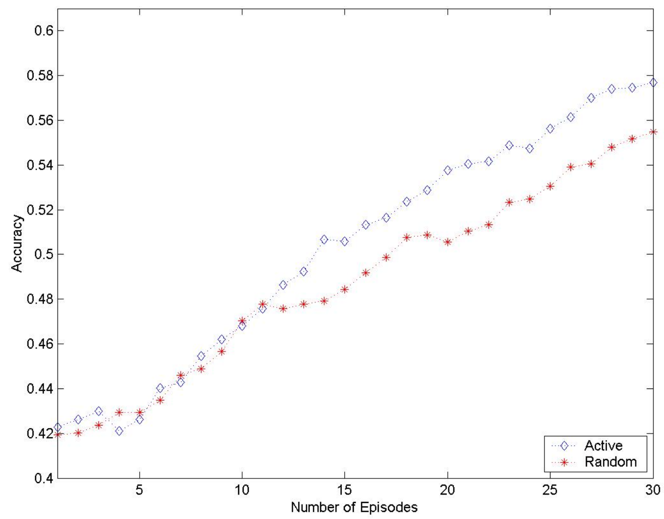

Figure 4.7: Model accuracy as number of trained episodes increases.

It appears that **the longer a user spends on a question, the more likely the question is to be answered incorrectly.**

6. The goal of collecting labeled examples is to use them to train the activity recognition model, thereby improving classification accuracy. We explored how much model improvement was gained by training on the episodes labeled during this experiment. Note that with perfect labeling of all 145 episodes we found that the maximum possible accuracy for the supervised learner was 73%. The starting accuracy for the model (i.e., the *baseline* accuracy) was 42.21%.

Accuracy improved as more episodes were labeled, however the rate of improvement differed between participants who labeled actively chosen episodes and those who labeled randomly chosen episodes. On average, the final accuracy after 30 episodes was higher for the actively chosen episodes (M=.53, SD=.036) than for the randomly chosen episodes (M=.50, SD=.034). This difference was significant, $t(20)$=2.17, p=.043. See Figure 4.7 for a graph of accuracy over both question sets as the number of labeled episodes grows from 1 to 30. These data indicate that **labels from the actively chosen question set improved model accuracy faster**

and pushed model accuracy higher than labels from the randomly chosen question set.

4.7 Discussion

In this section, we introduce several pieces of advice for future designs of context-aware recognition surveys.

- *Use "warm up" questions.* An active learning scheme is designed to present users with the most important questions first. However, our data suggested that a "learning curve" exists, with participants 1) taking longer to answer the first few questions, 2) performing worse on the first few answers, and 3) reporting lower confidence on the first few questions. We suggest opening with a few warm-up questions, so that participants will be more likely to answer the most important questions correctly.

- *Limit the number of displayed activities.* In the second experiment, the "multiple choice" aspect of our CARS included 27 possible activities. Such a large number of activities displayed at once made it difficult for participants to keep all possibilities in mind. In particular, we recommend that possible activities be segmented by room (as in our first experiment) and that similar activities be nested, so that the participant can "drill down" to greater levels of detail (e.g., Using Microwave drills down to Boiling Water in Microwave and Cooking Popcorn in Microwave, etc.).

- *Allow user to segment episodes.* Using automatic segmentation (i.e., episode recovery) will inevitably result in some number of badly segmented episodes. We found that using an active learning scheme decreased the number of these segments displayed to the user, however, at some point the user will likely be asked to label a question that is ill-segmented. In many cases, the boundaries between segments were obvious to the end user. We suggest allowing users to choose the start and finish of the episode, as well as the label and their confidence level.

- *Re-segment data on the fly.* In both reported experiments we segmented a number of activity episodes off-line, before participants were asked to perform labeling. When using HMMs for automatic segmentation, however, segmentation accuracy improves with a better model. Therefore, we suggest that after each new label is collected from the user (and the model subsequently updated/trained), the CARS program should perform a re-segmentation of the data, thereby improving segmentation accuracy and improving the likelihood that users will be presented with well-segmented, viable episodes of activity.

- *Model user abilities.* Labeling accuracy varied widely between individuals over both experiments. Significant accuracy might be gained by gauging the abilities of each user and presenting questions appropriate to the user's skill level. In this way, expert labelers could be utilized for difficult episodes and novice labelers could hone their skills on simpler fare. We suggest introducing redundancy by allowing multiple users to label the same episodes – following the QBC model, the most "challenging" episodes would be those with highest entropy between users. Possibly, the skill level of a labeler could be determined similarly, with highest skill accompanying those with lowest entropy between other labelers.

- *Display more timing information.* In our first experiment we displayed the total length of time that the episode required. In both experiments we displayed symbols that represented sensor readings chronologically, in the order that they were triggered by people acting inside the instrumented environment. We found that in some cases it could be useful to utilize *negative information*, in the form of gaps in time when the occupant did nothing. This information comes for free and is a key portion of many activities, most notably watching television.

- *Allow user to change scrolling speed.* Again, participants' abilities varied widely through both experiments. We suggest catering to user preferences by allowing control over the scrolling speed of symbols and captions.

4.8 Conclusion

In this chapter, we described an approach for data collection called the *context-aware* recognition survey. This approach uses contextual information collected by sensors to allow users to label episodes of activity in the home. CARS is desirable and practical because it does not require additional sensor infrastructure, does not interrupt the activities it collects data about, and allows anyone to label activities. We presented results from two experiments that indicate such an approach can be effective and at the same time remain completely unsupervised. We used our results to offer several "lessons learned" that could potentially help other researchers design a better CARS.

Chapter 5

Application to Activity Rating

Rating how well a routine activity is performed can be valuable in a variety of domains. In this chapter, we describe a general-purpose activity rating system built on the familiar hidden Markov model (HMM) framework.[1] We formalize the problem as MAP estimation in HMMs where the incoming trace needs repair. We present polynomial time algorithms for computing minimal repairs with maximal likelihood for HMMs, hidden semi-Markov models (HSMMs) and a form of HMMs incorporating partial temporal logic constraints. We present some results to show the promise of our approach.

5.1 Introduction

Rating how well a person performs a routine activity is a broadly useful capability with many applications: professors train medical students by rating their execution of established procedures, caregivers assess the well-being of their wards by rating how well they are able to perform activities of daily living, and managers and workflow experts identify poorly performed procedures that cause bottlenecks in a system. Although rating routine activity is certainly useful, as conventionally done it is also very expensive – each activity performance requires a dedicated human observer (often

[1]This chapter is a revised version of [120] and this work was performed in collaboration with researchers at Intel Research Seattle.

an expert). Many situations where gauging the performance of routine activities could be helpful are therefore either not rated at all, or rated in a cursory manner. Clearly, an opportunity exists for automated techniques to reduce the cost of rating. In this chapter, we explore methods for automatically rating performances of routine activities.

The basic classification task of rating, going from observations to scores, is amenable to a variety of standard approaches. Rating becomes challenging, however, if we wish to make it both *incrementally inexpensive* and *credible*. We define an incrementally inexpensive rater to be a rater in which the extra cost of rating a new activity is relatively low. The main determinant of cost is whether rating a new activity requires a custom classifier to be developed from scratch, or whether a generic classifier of some kind can be easily customized to the task. A credible rater is one that is both *relevant* and *transparent*. By relevant, we mean that the classification model for a particular rating task should reflect constraints on activity performance that are important to those using the rating. For example, a professor grading anesthesiology students performing an intubation may want to indicate what her notion of good performance is. By transparency, we mean that the system should be able to justify why it has assigned a particular rating. Ideally, the justification should be *constructive*, in that it should suggest how a low-rated performance may be altered to obtain a high-rated one.

Our techniques for rating activity routines are designed to satisfy the above requirements. To lower incremental cost, we choose a representation that is easily learned: all activities to be rated in our system are modeled by variants of hidden Markov models (HMMs). We intend that these models, especially given simple prior information, can be learned easily from training examples. More crucially, we formulate the justification for a rating relative to this model generically as the set of edits required on the trace generated by the rated activities; we therefore do not require special identification and modeling of errors and their causes. A fundamental weakness of these models is that they are first order, preventing them from capturing certain important correlations. We augment the Markov models with an intuitive constraint formalism (a small fragment of the temporal logic LTL [32]) that allows raters to explicitly state relevant constraints. Given these relevant and easy-to-construct models, we formulate rating as the likelihood of (possibly edited) observation sequences.

The core of this chapter consists of efficient algorithms to compute maximum likelihood paths of minimally edited versions of incoming observations with respect to various representations for activities, including HMMs, HSMMs and temporally constrained HMMs. The algorithms build on the dynamic programming technique used to great effect by the well-known Viterbi algorithm. We conduct a preliminary evaluation, demonstrating the promise of our technique.

5.2 Overview

In this section, we describe how we expect our system to be used, and we sketch how our system supports this usage model. Our goal is to develop a system that rates how well an elder performs day-to-day activities. Such a system is of great interest to the eldercare industry. In theory, caregivers will assess the elders' well-being by consulting ratings summaries and credible explanations of performance deficits. For example, the system may recognize that an elder is no longer able to prepare their daily bowl of soup, and report why (e.g., can't reach cabinet or difficulty holding spoon).

To end-users, our system represents activities as a set of steps. Each step has a duration and a set of observed actions performed, and is succeeded by other steps. For instance, the activity "making soup" for a particular elder may have the following steps: "preheat water," "open can," "mix and boil ingredients," "serve," and "clean up." The step "open can," may have an average duration of 45 seconds and contain the following actions: "use utensil drawer," "use can opener," "use can," and "use pantry door."

For concreteness, we will assume in what follows that we are using RFID-based [93] sensors that will directly sense the action of using particular objects. Therefore, all of our actions are of the form "use X" where X is some object. Inherently, our system requires that actions are observable by sensors. Given an activity trace (i.e., a trace of actions that constitutes a particular execution of an activity), our system provides a rating (e.g., *pass* or *fail*). If the grade is a *fail*, the system provides an alternate sequence of actions as close to the original as possible that would have elicited a *pass* grade (essentially a constructive justification of the grade). In more detail, use of the system proceeds as follows:

Learning the model: A human demonstrator performs the routine in an exemplary fashion. The system collects traces $Y_1, \ldots, Y_n$ of the routine. Each trace Y_i is a sequence of time-stamped observations $y_{i1}, \ldots, y_{im_i}$ of the demonstrator's actions. The traces are used to learn a dynamic stochastic model (either an HMM or an HSMM) with parameters λ. The hidden states $s_1, \ldots, s_N$ of the model correspond to the "activity steps" above, and are labelled $l_1, \ldots, l_N$ with the names of the step.

Adding global constraints: Typically, the first-order model learned in the previous step cannot capture important higher-order correlations. For instance, in a successful soup-making routine, the stove, if it is used, should eventually be turned off. The turning on would happen in the "preheat water" step, but the turning off may not happen until the end of the "serve" step. The human rater explicitly adds a set $\mathcal{C}$ of constraints on the sequence of hidden states or observations that specify these required higher-order correlations. In this case, a possible constraint would be of the form use("stove control knob") $\mathcal{E}$ use("stove control knob"), read as "a use of a stove control knob should eventually succeeded by a use of a stove control knob."

Learning rating thresholds: A human rater rates each trace Y_i with a rating $r_i \in \{pass, fail\}$. Let the *constrained MAP likelihood* of trace Y given model parameters λ and temporal constraints $\mathcal{C}$, $\hat{l}_Y = \text{CMAP}(M, Y, \mathcal{C})$, be the likelihood of the path $\hat{S}_Y$ with maximum a posteriori (MAP) likelihood given λ and Y that satisfies $\mathcal{C}$. We perform a simple thresholding computation to calculate the likelihood threshold L such that, given the classification function *R(l) = if* $l < L$ *then fail else pass*, $R(\hat{l}_{Y_i}) = r_i$ for as many of the Y_i as possible. Intuitively, the likelihood threshold L separates the passes from the fails.

Generating a rating and justification: Given the constrained model $(\lambda, \mathcal{C})$ and threshold L, the automated rater is ready for use. The person to be rated generates a trace $Y = y_1, \ldots, y_m$ to be rated automatically. The rater finds the constrained MAP likelihood $\hat{l}_Y$ and path $\hat{S}_Y = (\hat{s}_1, y_1), \ldots, (\hat{s}_m, y_m)$ for Y, and assigns it the rating $r = R(\hat{l}_Y)$. If $r = fail$, the rater attempts to produce a *repaired trace* trace $Y' = y'_1, \ldots, y'_m$ such that the edit distance between Y and Y' is as small as possible, and $\hat{l}_{Y'} > L$. In other words, Y' is the closest trace to T that passes. The rater offers r as the rating for the activity and, if appropriate, $\delta_{\hat{S}_Y, \hat{S}_{Y'}}$, the set of edits needed to transform $\hat{S}_Y$ into $\hat{S}_{Y'}$, as the justification for the rating.

As described above, our rating system employs two key non-standard pieces of machinery.

1. A method to compute the repaired observation trace T', that is a minimum edit distance from a given trace T with likelihood above threshold L.

2. A method to compute the constrained MAP likelihood function CMAP$(M, T, \mathcal{C})$.

5.3 Trace Repair for Hidden Markov Models

A hidden Markov model (HMM) $\lambda = (A, B, \pi)$ is a commonly used stochastic model for dynamic systems [96]. We formally pose the trace repair problem as a variation of estimating the most likely state sequence given a sequence of observations (classically solved via the Viterbi algorithm). An HMM is defined as follows. Let $Q_A = \{q_1, \ldots, q_N\}$ be the states of the process being modeled, and $O_B = \{o_1, \ldots, o_M\}$ the observation signals possibly generated by the process. We use meta-variables s_t and y_t to denote the states and observations respectively at time t. A_{ij} is the probability $p(s_{t+1} = q_j | s_t = q_i)$ of transitioning from state q_i at time t to q_j at time $t+1$ for any t; B_{ij} is the probability $p(y_t = o_j | s_t = q_i)$ of generating observation o_j when in state q_i (we write B_{iy_t} for B_{ij} such that $y_t = o_j$). The initial state distribution $\pi_i = p(s_0 = q_i)$.

5.3.1 The Repaired MAP Path Estimation Problem

We now formulate the problem of MAP path estimation given an observation sequence if we are allowed to first make a limited number of edits or "repairs" to the sequence. We begin by formalizing the notion of an edit. We then state the repaired MAP path estimation problem and present a variation of the Viterbi algorithm to solve it.

Let Y^N be the set of length-N strings of observations over some finite alphabet Y. Then $e^{k,N} = ((b_1, s_1), ...(b_N, s_N))$ is a *length-N k-Edit vector on Y^N*, with b_i boolean, s_i strings over Y, and $k = \sum_{1 \le i \le N}(b_i + |s_i|)$. For instance, $\hat{y}_1 =$ "cat" is a string in Y^3; $e_1^{4,3} = ((\text{false}, \text{"BB"}), (\text{false}, \text{""}), (\text{true}, \text{"R"}))$ is an edit vector on Y^3. Applying an edit vector e to string $\hat{y}_n = y_1 \dots y_n$, written $e(\hat{y})$ results in a new string $\hat{y'}$ obtained as follows. For $1 \le i \le n$, let if $e.b_i$ is true, then replace y_i with $\hat{y}'_i$, else replace y_i with $y_i e.s_i$ ($e.s_i$ appended to y_i). For example, $e_1^{4,3}(\hat{y}_1) =$ "cBBaR". A string $\hat{y}'$ is a *k-Edit* of another $\hat{y}$ if there exists edit vector $e^{k,N}$ such that $\hat{y}' = e^{k,N}(\hat{y})$.

We are now ready to specify the problem of MAP estimation with repairs:

Definition 1. (Repaired MAP Path Estimation Problem (RMAP)) *Given observation sequence $\hat{y}_T$, HMM $\lambda = (A, B, \pi)$ and edit distance K find observation sequence $\hat{y}'_{T'} = \dot{y}_1 \dots \dot{y}_{T'}$ that is a K-edit of $\hat{y}_T$ and path $\hat{s}_{T'} = \dot{s}_1, \dots, \dot{s}_{T'}$ maximizing $p(\hat{s}_{T'}, \hat{y}'_{T'})$ over all T', $\hat{s}_{T'}$ and $\hat{y}'_{T'}$.*

Before discussing our solution, we define string $\hat{y}'$ as the *(k,a)-edit* of string $\hat{y}_n$ if $\hat{y}' = e^{k,n}(\hat{y})$ for some $e^{k,n}$, $|e^{k,n}.s_n| \le a$, and additionally, $e^{k,n}.b_n$ if $a = 0$ and $|e^{k,n}.s_n| > 0$ if $a > 0$. The (k, a)-edit of a string requires its last character to be either preserved or replaced by at least one character, with at most a characters added. Edits compose as follows ($d_{\nu n} = n - \nu$; $a' = 1$ if $a \neq 0$, 0 otherwise; $(\nu, \alpha) <_k (n, a)$ if $\nu < n$ and $\alpha \le k$, or if $\nu = n$ and $\alpha < a$):

Lemma 1. (Edit Composition) *Let $Y_n^{ka\hat{y}}$ be the set of all (k, a)-edits of the n-prefix of string $\hat{y}$ over Y. Then $Y_n^{ka\hat{y}} = \{\hat{y}'y | y \in Y, \hat{y}' \in Y_\nu^{\kappa\alpha\hat{y}}, (\nu, \alpha) <_k (n, a), \kappa + d_{\nu n} + a' = k\}$.*

Table 5.1 specifies an algorithm (the k-Edit Viterbi (KEV) algorithm) to solve RMAP. KEV iterates over the T *original observations* in the incoming observation string. For each original observation, it iterates over possibilities for the K *added*

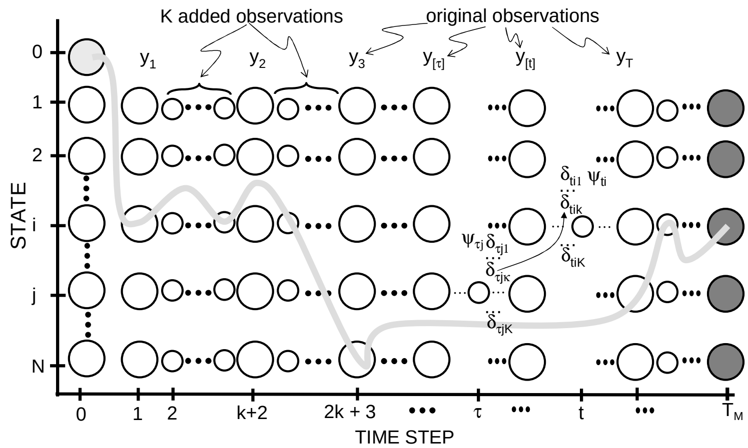

Figure 5.1: Trellis for k-Edits Viterbi on HMMs.

observations at that position, for a total of $TK + T$ iterations. At each iteration t corresponding to original obervation $[t] = t$ div $(K + 1) + 1$ and added observation $\#t = t \bmod (K+1)$, KEV computes the likelihood δ_{tik} of the most likely path ending in state i given an observation string that is a $(k, \#t)$-edit of $y_1 \ldots y_{[t]}$ over all such edit vectors; KEV also records as ψ_{tik} the penultimate state and edit in this path. Following the chain of ψ_{tik}'s back to the start state iteration gives the MAP repaired path.

The *trellis* of Figure 5.1 illustrates KEV. Columns of the trellis represent edits considered for inclusion into the final string. Large circles represent original observations and small ones represent adds. For technical reasons (to allow skipping the first original observation), we add a distinguished start state q_0 with new start probabilities $\pi'_0 = 1$, $A_{0i} = \pi_i$ and $A_{i0} = 0$ and add a column $(t = 0)$ processed in the initialization step. To allow skipping the last original observation with no adds, we add a column $(t = T_M = TK + T + 1)$. Rows represent possible hidden states.

The end result of the algorithm is a forward path (shown in light grey in Figure 5.1) through the trellis that, unlike in the conventional Viterbi algorithm, may jump between nodes in non-adjacent time slices. If the path jumps over the slice for an original observation y_i (where i is the position of the observation in the input string

Initialization:
$t = 0, \quad k = 1 \dots K, \quad 1 \leq i \leq N$

$$\delta_{t0k} = 1 \qquad \delta_{tik} = 0 \qquad \psi_{tik} = -1$$

Iteration:
$1 \leq t \leq TK + T, \quad k = a_t, a_t + 1, \dots, K$

$$(\psi_{tik})\delta_{tik} = \underset{\tau,j,\kappa \text{ s.t. } \kappa + a_t + d_{\tau t} = k}{(\arg)\max} \delta_{\tau j \kappa} B_{i y_{it}} A_{ji}$$

Termination:
$t = T_M = TK + T + 1$

$$(\psi_{tik})\delta_{tik} = \underset{\tau,j,\kappa \text{s.t. } \kappa + a_t + d_{\tau t} = k}{(\arg)\max} \delta_{\tau i \kappa} \ ;$$

$$i_M = \underset{1 \leq i \leq N}{\operatorname{argmax}} \, \delta_{tiK}$$

Backtracking:
$(t, i, k) = \psi_{T_M i_M K}; \qquad \text{while } t > 0,$

$$1) \quad (\dot{s}_t, \dot{y}_t) = (q_i, y_{it}) \qquad 2) \quad (t, i, k) \leftarrow \psi_{tik}$$

Table 5.1: The k-Edits Viterbi Algorithm for HMMs.

$\hat{y}$), we conclude that y_i was deleted from $\hat{y}$, otherwise not. Further, if the path passes through a sequence of added nodes with no intervening original node such that y_i is the first original observation to the left of the sequence, and the observations at these nodes are $y_{i_1}, \ldots, y_{i_n}$, we conclude that the string $y_{i_1} \ldots y_{i_n}$ was added at the i'th spot in the incoming string. The forward path is the required solution $\hat{s}^{T'}$, and the string of observations along the path is the edited string $\hat{y}^{T'}$.

The algorithm uses three intermediate variables, a_t, $d_{t\tau}$ and y_{it}. Variable $a_t = 1$ if $\#t \neq 0$ and 0 otherwise; $d_{\tau t} = [t] - [\tau]$, represents the number of deletes skipping original observations between τ and t; y_{it} is the observation considered when processing state i at slice t. Note that we only process original observations at time slices $1, K+1, 2K+1, ...$ In all other "added" slices, we need to propose the observed value to be added. A simple but inefficient approach would be to consider for each state, k-value and iteration t, every possible observable $o \in O_B$ as a candidate. In fact, we can consider a single observation instead of all $|O_B|$. The key insight is that, when processing state i in an added slice, it is sufficient to consider adding as observable the most likely observable in that state. Let S_N and Y_N be the sets of all length-N sequences of states and observables. Let $\dot{y}_i = \underset{1 \leq j \leq M}{\operatorname{argmax}} B_{ij}$. Let $\hat{s}q$ be the result of appending state q to sequence $\hat{s}$, and similarly for $\hat{y}y$. Then, for all $q_i \in Q_A$:

Lemma 2. $\underset{\hat{s},\hat{y}' \in S_N, Y_{N+1}}{max} p(\hat{s}q_i, \hat{y}') = \underset{\hat{s},\hat{y} \in S_N, Y_N}{max} p(\hat{s}q_i, \hat{y}\dot{y}_i)$

This follows from the fact that,

$$\begin{aligned} \max_{y_i} p(\hat{s}q_i, \hat{y}y_i) &= \max_{y_i} \pi_{s1} B_{s_1 y_1} \Big(\prod_{1 \leq i \leq N, s_i q_j \in \hat{s}_N} A_{ij} B_{j y_j} \Big) (A_{s_N i} B_{i y_i}) \\ &= \pi_{s1} \ldots A_{s_N i} \max_{y_i} B_{i y_i} = \pi_{s1} \ldots A_{s_N i} \dot{y}_i. \end{aligned}$$

Given this identity for the optimal observable to be added in state q_i, we set y_{it} to $\dot{y}_i$ if t is an "added" timeslice, and to $y_{[t]}$ otherwise.

We are now ready to establish the soundness of the KEV algorithm. Let S_n^i be the set of length-n sequence of states ending in state q_i. Let Y_n^{tik} be the set of length-n strings of observables that are $(k, \#t)$-edits of $y_1 \ldots y_{[t]}$.

Lemma 3. $(\psi_{tik})\delta_{tik} = \underset{n,\hat{s}\in S_n^i,\hat{y}\in Y_n^{tik}}{(arg)max} \; p(\hat{s},\hat{y})$

Proof sketch. Proof is by induction on t. We focus on the inductive case for δ. For ψ, replace "max" with "argmax".

$$\begin{aligned}
\delta_{tik} &= \max_{\tau,j,\kappa} A_{ji}B_{iy_{it}}\delta_{\tau j\kappa}\forall_{j,\tau,\kappa} \text{ s.t. } (\kappa + a_t + d_{\tau t}) = k \\
&= \text{(by the inductive hypothesis)} \\
&\quad \max_{\tau,j,\kappa}(A_{ji}B_{iy_{it}} \max_{n,\hat{s}\in S_n^j,\hat{y}\in Y_n^{\tau j\kappa}} p(\hat{s},\hat{y}))
\end{aligned}$$

Given s_n, s_{n+1} is independent of $\hat{s}^{n-1},\hat{y}^n$:

$$\begin{aligned}
A_{ji} &= p(s_{n+1} = q_i|\hat{s}, s_n = q_j, \hat{y}) \;\; \forall_{n,\hat{s}\in S_{n-1},\hat{y}\in Y_n^{\tau j\kappa}} \\
&= p(s_{n+1} = q_i|\hat{s},\hat{y}) \;\; \forall_{n,\hat{s}\in S_n^j,\hat{y}\in Y_n^{\tau j\kappa}}
\end{aligned}$$

Similarly, for y_{n+1} given s_{n+1}:

$$B_{iy_{it}} = p(y_{n+1} = y_{it}|\hat{s}, s_{n+1} = q_i, \hat{y}) \; \forall_{n,\hat{s}\in S_n^j,\hat{y}\in Y_n^{\tau j\kappa}}$$

Substituting for A_{ji} and $B_{iy_{it}}$ above, and using $p(A,B) = p(A|B)p(B)$ twice, we have, with $(\kappa + a_t + d_{\tau t}) = k$:

$$\delta_{tik} = \max_{\tau,j,\kappa,n,\hat{s}\in S_n^j,\hat{y}\in Y_n^{\tau j\kappa}} p(s_{n+1} = q_i, \hat{s}, y_{n+1} = y_{it}, \hat{y})$$

Lemma 2 ensures that maximizing over $\hat{y}y_{it}$ maximizes over all strings $\hat{y}y$. By lemma 1 maximizing over all $\hat{y}y$ with $\hat{y} \in Y_n^{\tau j\kappa}$ maximizes over $\hat{y} \in Y_{n+1}^{tik}$. Finally, $\forall_{1\le j\le N,\hat{s}\in S_n^j q_i}\hat{s} = \forall_{1\le i\le N,\hat{s}\in S_{n+1}^i}\hat{s}$. Modifying the previous equation to reflect these insights:

$$\delta_{tik} = \max_{n,\hat{s}\in S_{n+1}^i,\hat{y}\in Y_{n+1}^{tik}} p(\hat{s},\hat{y}) = \max_{n,\hat{s}\in S_n^i,\hat{y}\in Y_n^{tik}} p(\hat{s},\hat{y})$$

□

The soundness of KEV follows in a straightforward way from the above lemma. Further, given that the trellis has $O(TKN)$ nodes, that at each node we compute

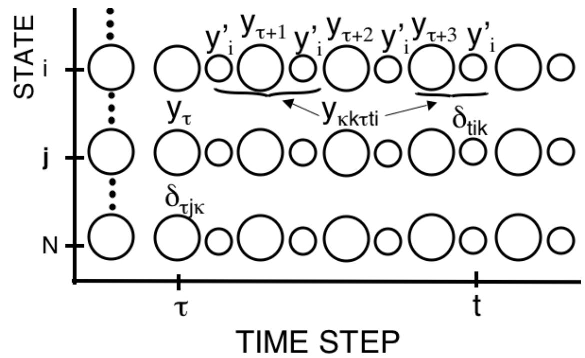

Figure 5.2: Trellis for k-Edits Viterbi on HSMMs.

$O(K)$ δ and ψ values, and we consult $O(NK)$ preceding data values to do so, the complexity of KEV as a whole is $O(TN^2K^3)$.

5.4 Trace Repair for Hidden Semi-Markov Models

A hidden semi-Markov model (HSMM) [88] $\lambda = (A, B, D, \pi)$ is identical to an HMM except for the *duration distribution* D. Where an HMM generates a single observation according to B on each visit to a state s, the HSMM generates l independent observations from B on each visit, where l is drawn according to $D_{sl} = p(l|s)$. The added flexibility is useful when modeling human activities, since the duration of stay in a state is restricted to be geometric (and therefore biased to small values) in HMMs. In what follows, we assume that D is over a finite set (of size $|D|$) of durations, where the longest duration is L steps.

The RMAP problem is: given HSMM (A, B, D, π), observations $\hat{y}^T$ and limit K, find $\underset{t, \hat{s} \in Q_A^t, \hat{y} \in O_B^t, \hat{y} \text{ } k\text{-Edit of } \hat{y}^T}{\operatorname{argmax}} p(\hat{s}, \hat{y})$.

Table 5.2 specifies a variant of KEV to solve the problem, and Figure 5.2 shows a trellis for this algorithm. The trellis is identical to that used by KEV (we represent k added nodes with a single small circle), only its use is different. We focus on how δ_{tik} is calculated. At each timestep t, state i and edit distance k, as with KEV, we iterate over previous timesteps, states and edit distances τ, j and κ. However, this time instead of discarding the observations in the intervening timesteps, we seek their

Initialization:
$t = 0, \quad k = 1 \ldots K, \quad 1 \le i \le N$

$$\delta_{t0k} = 1 \qquad\qquad \delta_{tik} = 0 \qquad\qquad \psi_{tik} = -1$$

Iteration:
$1 \le t \le TK + T, \quad k = a_t, a_t + 1, \ldots, t$

$$(\psi_{tik})\delta_{tik} = \underset{1\le\tau\le t, j, \kappa, y_{\kappa k\tau ti}}{(\arg)\max} \ \delta_{\tau j\kappa} p(y_{\kappa k\tau ti}) A_{ji} D_{i|y_{\kappa k\tau ti}|}$$

Termination:
$t = T_M = TK + T + 1$

$$(\psi_{tik})\delta_{tik} = \underset{\tau, j, \kappa \text{s.t. } \kappa + a_t + d_{\tau t} = k}{(\arg)\max} \delta_{\tau i\kappa} \ ;$$
$$i_M = \underset{1\le i\le N}{\operatorname{argmax}} \ \delta_{tiK}$$

Backtracking:
$(t, i, k) = \psi_{T_M i_M K}; \qquad \text{while } t > 0,$

$$1) \quad (\dot{s}_t, \dot{y}_t) = (q_i, y_{it}) \qquad 2) \quad (t, i, k) \leftarrow \psi_{tik}$$

Table 5.2: The k-Edits Viterbi Algorithm for HSMMs.

sub-sequence $y_{\kappa k\tau ti}$. We assume that step t only ends a stay in state q_i that begins immediately after the stay in q_j that ended in step τ. If $e_{\tau t}$ is the number of edits in $y_{\kappa k\tau ti}$ (added nodes included + original nodes ignored), we require $\kappa + a_t + e_{\tau t} = k$. The problem of maximizing the likelihood of the path ending at (i, t) then reduces to the problem of finding $y_{\kappa k\tau ti}$ maximizing $p(y_{\kappa k\tau ti})D_{i|y_{\kappa k\tau ti}|}$.

We find this maximum by iterating through durations l in D_i; for each l, we iterate through predecessors (τ, j, κ) of (t, i), finding a sequence $y_{\kappa k\tau ti}$ of length l with the highest probability; we keep a running tally of the maximum $p(y_{\kappa k\tau ti})D_{il}$. Finding $y_{\kappa k\tau ti}$ reduces to identifying N_A added nodes (to include in $y_{\kappa k\tau ti}$) and N_O original nodes (to ignore), such that $N_A + N_O = k - \kappa - a_t$ (to satisfy the k-Edit criterion), and $N_A + (([t] - [\tau]) - N_O) = l$ (to satisfy the duration constraint). The two equations fix N_A and N_O. Since all the added nodes have the same probability $\dot{y}_i = p(\hat{o}_i|q_i)$, it doesn't matter which particular N_A we pick. On the other hand, we pick the N_O original nodes with lowest probability of observation for exclusion; this can be done by sorting the original nodes in $O(T\log T)$ time *offline*, with $O(L)$ access during execution. Once the sequence of nodes is picked to get $y_{\kappa k\tau ti}$, we simply multiply their observation and transition probabilties together to get $p(y_{\kappa k\tau ti})$, a process that takes l operations, since $|y_{\kappa k\tau ti}| = l = O(L)$.

Given $O(TNK)$ trellis nodes, computing $O(K)$ δ and ψ values at each node, consulting $O(|D|NK)$ preceding values for each value, and spending $O(L)$ for each preceding value considered, the entire algorithm takes $O(TN^2|D|LK^3)$ steps. Note that in the (fairly) common case that D and L are unbounded, this running time becomes $O(T^3N^2K^3)$.

5.5 Trace Repair for Constrained HMMs

We define a *temporally constrained HMM (TCHMM)* as $\lambda = (A, B, \mathcal{C}, \pi)$, where $\mathcal{C}$ is a *temporal constraint* of the form $\phi_1\mathcal{E}\phi_2\mathcal{E}\ldots\mathcal{E}\phi_{|\mathcal{C}|}$. The ϕ_i are propositional boolean formulas over state labels l and observations y: $\phi ::= \text{state}(l)|\text{obs}(y)|\phi\wedge\phi|\neg\phi$. Path suffix $s_i \ldots s_T$ and observations $y_i \ldots y_T$ satisfy the constraint suffix $C_j = \phi_j \ldots \phi_W$ if for any $k \geq j$, $\phi_j(s_k, y_k)$ implies that $(s_{k+1} \ldots s_T, y_{k+1} \ldots y_T)$ satisfy C_{j+1}, written $(s_{k+1} \ldots s_T, y_{k+1} \ldots y_T) \vdash C_j$. Intuitively if one formula in the constraint sequence is true w.r.t. the head of the state/observation sequences, then the formulas that follow must also eventually be true in their specified order later in the sequences. The constraint $(\text{state(COOK)} \wedge \text{obs(oil)})\mathcal{E}(\text{state(WASH)} \wedge \text{obs(soap)})$ could, for instance capture the constraint that if oil is used in the cooking step of making dinner, soap should be used in the eventual required washing step.

The RMAP problem may now be reformulated as given TCHMM with constraints $\mathcal{C}$, observations $\hat{y}^T$ and limit K, find $SY = \underset{t,\hat{s}\in Q_A^t,\hat{y}\in O_B^t,\hat{y}\ k\text{-Edit of }\hat{y}^T}{\text{argmax}} \phi(\hat{s},\hat{y})$ such that $SY \vdash \mathcal{C}$.

Our solution for RMAP estimation is restricted to formulas of the form $\phi ::= \text{state}(l)|\phi\wedge\phi|\neg\phi$ (we disallow dependences on observables). A small modification to the KEV algorithm enables polynomial time solution of this problem. We use the same trellis as in KEV. For each timestep t, state i and edit distance k, we also now maintain an additional $|\mathcal{C}|$-vector. An element δ_{tikm} with $0 \leq m < |\mathcal{C}|$ represents the likelihood of the MAP path ending at state i in time slice t with $(k, \#t)$ edits that still requires constraint suffix C_{m+1} to be satisfied (except δ_{tik0}, which has no outstanding constraints to be satisfied). This likelihood can be computed compositionally from $\delta_{\tau j \kappa \mu}$, with $\tau < t$ and (κ, μ) pointwise $\leq (k, m)$ in $O(TN^2k^3|C|^2P)$ steps, where formulas ϕ_i can be evaluated in $O(P)$ steps (where P is the size of the formulas).

Even MAP estimation (without trace perturbation) for TCHMM's has apparently neither been formulated nor solved previously, although it is potentially quite powerful. For instance, the constrained inference work of Culotta *et. al.* [37] is a special case of TCHMM k-Edit MAP estimation (with $k = 0$, and $\mathcal{C} = \text{state}(q_0)\mathcal{E}\text{state}(q_i)$). MAP estimation is a special case of RMAP estimation with $k = 0$. Our variant of KEV above therefore performs MAP estimation. Interestingly with $k = 0$, we can allow the more general version of formulas ϕ and still retain the fast running time. It is open how general $\mathcal{C}$ can be while remaining tractable. For instance, our constraints can be viewed as a fragment of Linear Temporal Logic (LTL) [32]. It is interesting to consider larger fragments as candidates.

5.6 Evaluation

How does model choice affect advice? The k-Edits Viterbi algorithm dispenses advice based on the parameters of its activity models. The credibility of this advice will suffer from any differences between these models and the reality they represent. In order to illustrate this point, we conducted two experiments over three different activity models. The two experiments compare a regular HMM and a time-sensitive HSMM, and a regular HSMM with an HSMM that has temporal logic constraints, respectively.

First, we compare the output of HMMs versus HSMMs on three activity traces from different activity models (see the top row of Figure 5.3). Each activity trace was intentionally made incorrect: for making tea, the preparation step was hurried; for making a sandwich, not enough ingredients were collected; and for grooming, brushing

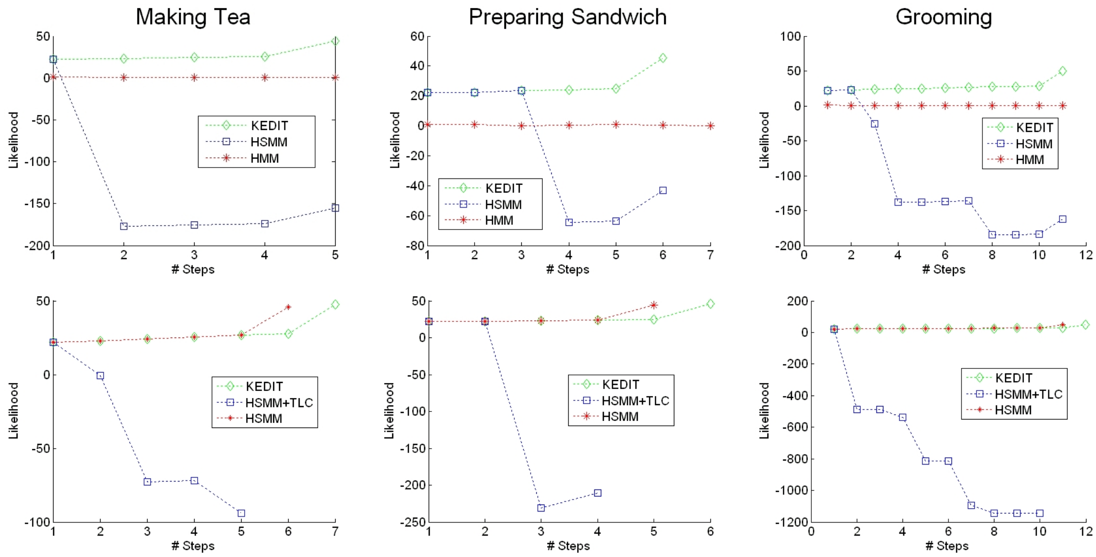

Figure 5.3: HMMs vs. HSMMs (top) and HSMMs vs. TCHMMs (bottom).

teeth and combing hair were performed too rapidly. In the top row of Figure 5.3, we plotted the maximum likelihood values at each step of the activity traces (where a “step” is considered to be a state transition). HMMs fail to detect any problem, exhibiting high likelihood. However, HSMM likelihoods plummet, due to sensitivity to the amount of time spent in each state. The KEDIT trace correctly adds the proper number of observations to each state, resulting in a high likelihood.

Second, we compare the output of HSMMs with and without temporal logic constraints (TLCs) (see the bottom row of Figure 5.3). Again, we intentionally chose incorrect sequences for the three activities: for making tea, the stove is turned on but never turned off; for preparing a sandwich, the refrigerator door is opened and never closed; and for grooming, the sink water is turned on and never turned back off. In the top row of Figure 5.3, we plotted the maximum likelihood values at each step of the activity traces. Regular HSMMs fail to detect any problem, reporting high likelihood. HSMMs with TLCs report low likelihood, because they are only allowed to consider state-transitions which satisfy all constraints. In these traces, constraints are broken and alternate, low-likelihood, paths must be considered. The KEDIT trace correctly adds the necessary steps (i.e., turn off stove, shut refrigerator, and turn off

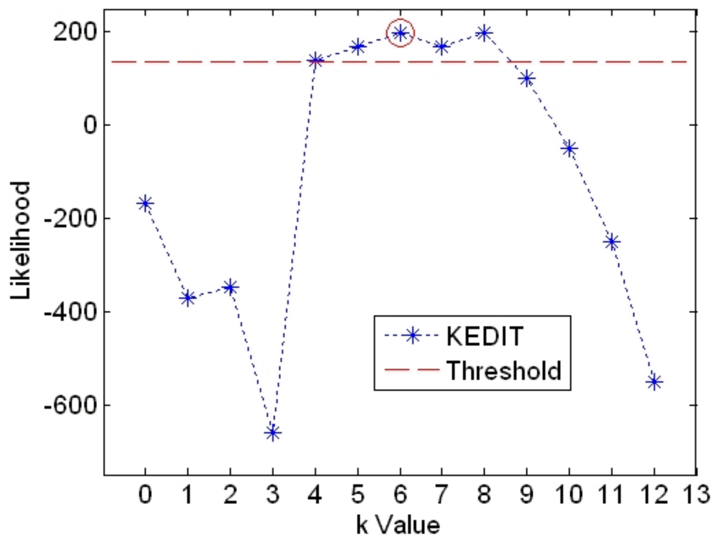

Figure 5.4: The likelihood of KEDIT traces as k increases.

sink), resulting in high likelihood.

How does the rating change as k increases? The k-Edits Viterbi algorithm provides advice for up to K edits. Ideally, we desire a trace that is above the likelihood threshold with the minimum number of edits. One method is to incrementally increase k until the threshold is exceeded. For this reason, we are interested in how the likelihood changes as k increases.

In this experiment we ran k-Edits Viterbi for HSMMs on an empty trace of the "making tea" activity. In Figure 5.4 we plotted the overall likelihood of each trace as the number of possible edits was increased. The dashed line is a threshold showing the likelihood of an acceptable "good" trace. Obviously, the original empty sequence had low likelihood. As k was increased from one to three, the algorithm was forced to assemble partially complete activity traces which had even lower overall likelihood. When $k = 4$ the algorithm formed a complete trace and met the threshold. As k increased further, the algorithm tweaked the sequence for a slightly higher likelihood. The most likely possible path was reached at $k = 6$. Afterwards, we see an "odd-even" effect as the algorithm is forced to add new (less likely) observations, and then opportunistically delete other observations. For $k \geq 9$, the likelihood drops as the algorithm performs too many modifications to the trace and is unable to reach the optimal solution.

How intuitive is the advice? We now examine the advice dispensed by k-Edits Viterbi in several scenarios. We ran the algorithm on activity traces that had the following problems: restarted the activity, got two steps out of order, performed a step too quickly, and missed or sped through several non-consecutive steps. All traces are from the "making tea" activity and likelihoods are reported using the optimal number of edits (i.e., k value).

The beginning steps of the next trace were performed twice (i.e., a "start" and a "restart"). The algorithm finds the maximum likelihood solution by deleting extra observations. However, the algorithm did not delete the entire start or restart, but decided to "pick and choose" among the two, keeping the best observations of both. In contrast, our intuition would be to advise the user to keep either the start or the restart. Similarly, in a trace in which two steps were performed out of order, the algorithm deletes one of the mis-ordered steps and inserts new steps in the correct position. We found this to be less intuitive than simply telling the user to switch the two steps to the correct order.

In the next trace, one step was performed too quickly; the "preparation step" only generated one observation, when it should have generated at least two. The algorithm suggested new observations that corrected the amount of time is spent in the state. However, the algorithm will always suggest the most likely observation from the state, because this maximizes the overall likelihood. We found this suggestion strategy to be non-intuitive (although mathematically optimal), however, it became a non-issue for models in which observations were spread across multiple states.

In the last trace several non-contiguous steps were missed entirely or performed too quickly. As k increased, the algorithm first chose to insert states that had been missed entirely, and then to add more observations to states that had been visited too briefly. In other words, the algorithm advises the user to at least visit each step of the activity before it advises how to perfect each step. This "top-down" approach fits with our intuition of how advice should be given.

5.7 Conclusions

In this chapter, we described the credible activity rating problem. We introduced the k-Edits Viterbi algorithm and showed that given model parameters and an activity trace it can provide optimally repaired traces with from zero to k edits. We improved the algorithm by incorporating high-level temporal logic constraints. Finally, we evaluated the strengths and limitations of the algorithm on data from three activity models.

Chapter 6

Conclusion

In this thesis, we focused on the area of automatic health monitoring, in which health information is automatically collected with the help of sensors and learning algorithms and distributed to caregivers. The work in this thesis – both qualitative studies and quantitative experiments – contributes to the present and future of automatic health monitoring in the following ways:

- In-depth knowledge of current practices in in-home health monitoring could inform the development of future technologies, ultimately helping an increased portion of the growing elderly population to live safely and independently in their own homes.
- Algorithms using sensors common to security systems could relatively instantly introduce ubiquitous computing services to thousands of corporate and residential buildings, possibly changing the way our society lives and works.

In this thesis work, we conducted a nationwide study of the professionals who currently perform in-home health assessment and identified promising areas for technological innovation. We posed the simultaneous tracking and activity recognition problem and offered a particle filter-based solution. We invented a novel data collection technique to help meet the practical need for labeled training examples. Finally, we approached an important application area called activity rating, and provided a unique algorithm which can pinpoint and explain irregularities in routine behavior. In the next section, we review each of these contributions in more detail.

6.1 Summary

In this section, we briefly summarize the major findings of this thesis work:

6.1.1 The Activities of Daily Living Study

In this study, we interviewed and distributed questionnaires to nearly one hundred professionals who routinely conduct in-home health assessments. Our findings were broadly applicable to people tracking, activity recognition, and the study of care networks for elders.

- We enumerated a "top ten" list of the most valuable activities of daily living.
- We identified gaps between privacy and perceived privacy constraints, e.g., motion detectors are okay as part of a home security system, but not alone.
- We described the home environment, including the number of occupants, presence of pets, public and private locations, and usage of assistive technology.

6.1.2 Simultaneous Tracking & Activity Recognition

In this work, we sought to automatically collect the information most important to automatic health assessment, including: location, locomotion, activities of daily living, extended activities of daily living, and instrumented activities of daily living. We defined the simultaneous tracking and activity recognition problem, whose solution provides this information. We utilized Bayes filters and particle filters to provide location estimation and activity recognition for multiple occupants and evaluated our approach on three different data sets.

- We showed that a Bayes filter which leverages location estimation to improve activity recognition can effectively recognize several activities of daily living.

- For tracking multiple occupants, we demonstrate a Rao-Blackwellised particle filter approach that can efficiently perform data association on information from anonymous sensors.

- We explored a neglected set of pre-existing sensors that are non-invasive, cheap, and easy to install and maintain, to introduce cost-effective automatic health monitoring.

6.1.3 The Context-Aware Recognition Survey

In this work, we built a completely unsupervised data collection technique in which information collected from an existing sensor infrastructure is placed into a game-like computer program where users are able to label anonymous episodes of activity.

- We described a novel data collection technique which can be used alone or in concert with existing methods to provide labeled training examples without interrupting the daily routines it is designed to learn about.

- We showed that participants were able to successfully label episodes of activity, and we narrowed down the conditions under which accuracy was optimal.

- We provided a series of "lessons learned" to help future researchers design a better data collection vehicle.

6.1.4 The k-Edits Viterbi Algorithm

In this work, we derived a new algorithm which is a more general version of the well-known Viterbi algorithm to provide automatic activity rating. Given HMM model parameters and an input trace, our polynomial time approach provides the maximum a posteriori likelihood sensor trace with up to k edits.

- We define the credible activity rating problem, identifying key components of a good rater, including incremental inexpensiveness, credibility, and justifiability.

- We extended the algorithm to be compatible with time sensitive hidden semi-Markov models.
- We further extended the algorithm to be compatible with high-level temporal logic constraints.

Appendix A

ADL Study Phase # 1: Questionnaire

SHARP: Activities of Daily Living Study

NOTE: *This questionnaire includes a number of questions about you. Please answer them as honestly and accurately as you can, as they will help us understand what you do. Thank you!*

Intel Research Use Only
Part. ID: ________________

1. Your **gender**: ❐ Male ❐ Female

2. Your **age**: __________

3. Your **job title**: ____

4. How **many years** have you **worked at this type of job**? ______________

5. In **which city/county** do **you live**?____

6. Which **cities/counties** do you **regularly visit** for your job?__________________________

 ____

 ____

7. How large is your **average case load**? ______ **your smallest?** ______ **your largest?** ______

8. How many **visits** do you usually make **per day?** _____ **per week?** _____ **per month?** _____

9. How often do you **use or do** the following during the course of your work?

eMail:	[frequently]	[occasionally]	[rarely]	[never]		
Internet:	[frequently]	[occasionally]	[rarely]	[never]		
Voice mail:	[frequently]	[occasionally]	[rarely]	[never]		
Electronic calendar/schedule:	[frequently]	[occasionally]	[rarely]	[never]		[what's that?]
Purchase from web sites:	[frequently]	[occasionally]	[rarely]	[never]		[what's that?]
Instant messaging:	[frequently]	[occasionally]	[rarely]	[never]		[what's that?]
Text messaging:	[frequently]	[occasionally]	[rarely]	[never]		[what's that?]
OnStar:	[frequently]	[occasionally]	[rarely]	[never]		[what's that?]
Computer or laptop:	[frequently]	[occasionally]	[rarely]	[never]	[don't have one]	
Answering machine:	[frequently]	[occasionally]	[rarely]	[never]	[don't have one]	
Cell phone:	[frequently]	[occasionally]	[rarely]	[never]	[don't have one]	
PDA—e.g., Palm Pilot:	[frequently]	[occasionally]	[rarely]	[never]	[don't have one]	[what's that?]
MP3 player:	[frequently]	[occasionally]	[rarely]	[never]	[don't have one]	[what's that?]

Thank you!

Figure A.1: Questionnaire from ADL study phase #1.

Appendix B

ADL Study Phase # 1: Questionnaire Results

Personal Information

Participant IDs:	01 - 05		
Gender:	Female=100%, Male=0%		
Age:	M=42.8	SD=7.16	RANGE=32-52
City of residence:	Tulsa		

Work information:

Job title:	Case Manager		
Years worked at this type of job:	M=5.2	SD=3.11	RANGE=2-10
Cities/Counties visited during job:	Tulsa, Osage		
Average case load:	M=26.5	SD=27.14	RANGE=3-50
Smallest case load:	M=24	SD=24.25	RANGE=3-45
Largest case load:	M=54.6	SD=34.49	RANGE=5-102
Visits made per month:	M=25	SD=14.47	RANGE=3-40

Appliances and electronics on the job:

How often they use the following devices or systems…

eMail:	Frequently=80%	Occasionally=0%	Rarely=20%	Never=0%
Internet:	Frequently=40%	Occasionally=20%	Rarely=20%	Never=20%
Voice mail:	Frequently=100%	Occasionally=0%	Rarely=0%	Never=0%
Electronic calendar:	Frequently=20%	Occasionally=40%	Rarely=40%	Never=0%
Online purchasing:	Frequently=0%	Occasionally=0%	Rarely=20%	Never=80%
Instant messaging:	Frequently=0%	Occasionally=0%	Rarely=60%	Never=40%
Text messaging:	Frequently=20%	Occasionally=20%	Rarely=20%	Never=40%
OnStar:	Frequently=0%	Occasionally=0%	Rarely=0%	Never=100%
Computer or laptop:	Frequently=100%	Occasionally=0%	Rarely=0%	Never=0%
Answering machine:	Frequently=80%	Occasionally=0%	Rarely=0%	Never=0%
Cell phone:	Frequently=80%	Occasionally=20%	Rarely=0%	Never=0%
PDA:	Frequently=0%	Occasionally=0%	Rarely=0%	Never=0%
	Don't have one=80%			
MP3 player:	Frequently=0%	Occasionally=20%	Rarely=0%	Never=20%
	Don't have one=60%			

Figure B.1: Responses to questionnaire from ADL study phase #1.

Appendix C

ADL Study Phase # 1: Interview Guide

General Duties

1. Please describe your overall job.
   [Probe: How many consumers do you have? What do you do in the field vs. in the office? What do you do for your consumers? What decisions do you make for your consumers?]

   a. What are your main responsibilities?
   b. About how much time do you spend on each part of your job?
   c. What is your favorite part of the job?
   d. What is the hardest part of your job?

Visiting a Consumer

2. Please walk me through a typical visit with a consumer.
   [Probe: What time do you go? What do you bring? What happens after you step in the front door?]

   a. What is your biggest priority during the visit?
   b. How long does a visit usually take?
   c. Please describe the range of problems that your consumers have.
   d. Please describe which ADLs you pay attention to.
   e. What do you do if no one is home? How often does this happen?

3. During a visit, how do you get information about ADLs?
   [Probe: What do you say? Who do you talk to?]

   a. What sort of questions do you ask the consumer?
   b. What sort of questions do you ask other people?
   c. What other clues do you pay attention to?
   d. Where do you typically go in the house?
   e. What is the most challenging part of collecting information about ADLs?

4. How do you feel about making visits?
   [Probe: Are visits worthwhile? Would you rather be at the office or in the field? How do you feel about seeing your consumers?]

   a. What is the hardest part of a visit?
   b. How do consumers typically feel about your visits?

5. How do you prepare for a visit with a consumer?
   [Probe: Do you do any paperwork? Do you consult with other nurses or physicians?]

   a. What do you prepare?
   b. How long does your preparation take?
   c. Who do you interact with while preparing for a visit?
   d. Do you have contact with the consumer between visits?

6. How do you schedule consumer visits?
   [Probe: Do you decide or does someone else? How far in advance? What factors are taken into consideration? Is it subject to change? Who actually schedules the visit? Who hires the case manager?]

Figure C.1: Interview guide from ADL study phase #1 – page 1.

Flow of Information

7. Please describe the ADL forms that you fill out.
   [Probe: What is covered? Do you use a new one every time? How much does the information change between visits?]

   a. Where do the forms come from? Who designed them?
   b. Are forms the same for every consumer? If no, how do they differ?
   c. Where are you when you fill them out?
   d. How often do you fill them out?
   e. About how much time does it take to fill them out? How does this vary?
   f. Which portions are most/lease time consuming? Why?
   g. Which portions are most/least important? Why?
   h. Which portions are hardest/easiest to fill out? Why?

8. After a visit is over, what happens to the forms you collected?
   [Probe: Where do they end up? Do other people see them?]

   a. Who else uses the forms?
   b. Are copies made?
   c. Where are the forms sent to?
   d. Where are the forms stored?
   e. How long are the forms stored?
   f. Who has access to the forms?
   g. What do they use the forms for?
   h. How often are the forms used?
   i. Who is responsible for the forms?
   j. Do you ever refer to your old forms? The old forms of others?

Social Network (who's who)

9. Outside of the agency, who else is involved in taking care of a consumer?

   a. What is their level of involvement?
   b. What do these other people do for the consumer?
   c. How much time do they spend doing it?
   d. In what ways do you interact with these other people?
   e. How often do you interact with them?
   f. How do you feel about interacting with them?

10. Do your consumers ever interact with each other?

Figure C.2: Interview guide from ADL study phase #1 – page 2.

New Technology

11. Which ADLs would be most/least important to know?
    [Probe: If you could have one ADL reported, which would it be?]

    a. Why is this ADL more important than the others?
    b. Why is this ADL less important than the others?

12. How would you feel about a tool that could automatically inform you of which ADLs are happening in a consumer's home?

    a. How do you feel about the idea in general?
    b. What would be the best part of having this information?
    c. Do you see any drawbacks?

13. How might this technology change the way you do your job?
    [Probe: How would your scheduling methods change? Your visits?]

    a. How would you want to receive this information?

14. How do you think this technology would affect the privacy of the consumer?

    a. Who do you think should have access to this information?

Wrap-up

15. Is there anything that I have forgotten to ask that you feel is important?

Thank you!

Figure C.3: Interview guide from ADL study phase #1 – page 3.

Appendix D

STAR Experiment # 1: Motion Models

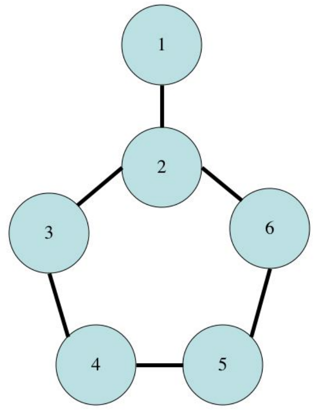

Figure D.1: Physical layout for STAR experiment # 1.

D.1 Physical Layout

D.2 Models

$$\begin{array}{c|cccccc|} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 1 \\ 3 & 0 & 1 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 1 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 1 & 0 & 0 \\ 6 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}$$

Table D.1: SAME model.

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	0	0	0	0	0	1
3	0	1	0	0	0	0
4	0	0	1	0	0	0
5	0	0	0	1	0	0
6	0	0	0	0	1	0

Table D.2: OPPOSITE model, occupant A.

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	0	0	1	0	0	0
3	0	0	0	1	0	0
4	0	0	0	0	1	0
5	0	0	0	0	0	1
6	0	1	0	0	0	0

Table D.3: OPPOSITE model, occupant B.

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	0	0.25	0	0	0	0.75
3	0	0.75	0.25	0	0	0
4	0	0	0.75	0.25	0	0
5	0	0	0	0.75	0.25	0
6	0	0	0	0	0.75	0.25

Table D.4: MIDDLE model.

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	0	0.33	0.33	0	0	0.33
3	0	0.33	0.33	0.33	0	0
4	0	0	0.33	0.33	0.33	0
5	0	0	0	0.33	0.33	0.33
6	0	0.33	0	0	0.33	0.33

Table D.5: UNIFORM model

Appendix E

STAR Experiment # 1: Small House Layout

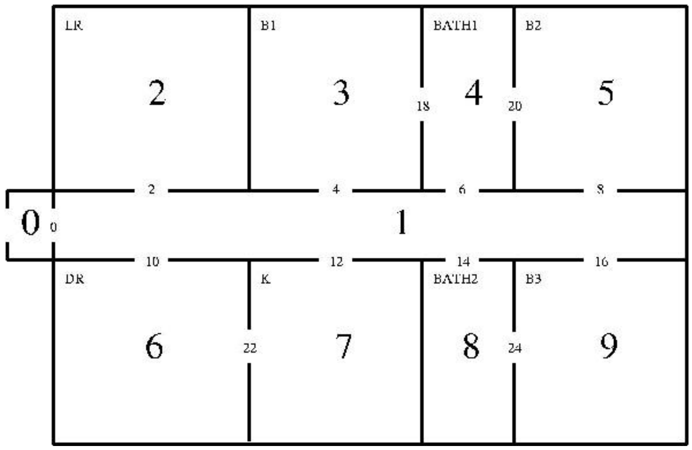

Figure E.1: Physical layout of STAR experiment # 1.

E.1 Physical Layout

E.2 Models

	0	1	2	3	4	5	6	7	8	9
0	0.1	0.9	0	0	0	0	0	0	0	0
1	0.1	0.04	0.2	0.04	0.04	0.04	0.1	0.05	0.19	0.2
2	0	0.05	0.95	0	0	0	0	0	0	0
3	0	0.85	0	0.1	0.05	0	0	0	0	0
4	0	0.7	0	0.1	0.1	0.1	0	0	0	0
5	0	0.6	0	0	0.1	0.3	0	0	0	0
6	0	0.1	0	0	0	0	0.7	0.2	0	0
7	0	0.6	0	0	0	0	0.3	0.1	0	0
8	0	0.1	0	0	0	0	0	0	0.7	0.2
9	0	0.1	0	0	0	0	0	0	0.2	0.7

Table E.1: Occupant A.

	0	1	2	3	4	5	6	7	8	9
0	0.1	0.9	0	0	0	0	0	0	0	0
1	0.1	0.04	0.1	0.04	0.18	0.18	0.1	0.2	0.04	0.02
2	0	0.4	0.6	0	0	0	0	0	0	0
3	0	0.6	0	0.1	0.3	0	0	0	0	0
4	0	0.1	0	0.05	0.7	0.15	0	0	0	0
5	0	0.1	0	0	0.2	0.7	0	0	0	0
6	0	0.1	0	0	0	0	0.6	0.3	0	0
7	0	0.1	0	0	0	0	0.2	0.7	0	0
8	0	0.7	0	0	0	0	0	0	0.2	0.1
9	0	0.6	0	0	0	0	0	0	0.1	0.3

Table E.2: Occupant B.

	0	1	2	3	4	5	6	7	8	9
0	0.1	0.9	0	0	0	0	0	0	0	0
1	0.15	0.02	0.2	0.2	0.17	0.02	0.1	0.1	0.02	0.02
2	0	0.2	0.8	0	0	0	0	0	0	0
3	0	0.05	0	0.8	0.15	0	0	0	0	0
4	0	0.05	0	0.2	0.7	0.05	0	0	0	0
5	0	0.3	0	0	0.6	0.1	0	0	0	0
6	0	0.1	0	0	0	0	0.6	0.3	0	0
7	0	0.1	0	0	0	0	0.4	0.5	0	0
8	0	0.9	0	0	0	0	0	0	0.05	0.05
9	0	0.9	0	0	0	0	0	0	0.05	0.05

Table E.3: Occupant C.

	0	1	2	3	4	5	6	7	8	9
0	0.1	0.9	0	0	0	0	0	0	0	0
1	0.15	0.1	0.3	0.1	0.1	0.03	0.04	0.1	0.04	0.04
2	0	0.05	0.95	0	0	0	0	0	0	0
3	0	0.1	0	0.8	0.1	0	0	0	0	0
4	0	0.1	0	0.3	0.5	0.1	0	0	0	0
5	0	0.7	0	0	0.2	0.1	0	0	0	0
6	0	0.3	0	0	0	0	0.6	0.1	0	0
7	0	0.3	0	0	0	0	0.5	0.2	0	0
8	0	0.85	0	0	0	0	0	0	0.1	0.05
9	0	0.85	0	0	0	0	0	0	0.1	0.05

Table E.4: Occupant D.

	0	1	2	3	4	5	6	7	8	9
0	0.1	0.9	0	0	0	0	0	0	0	0
1	0.2	0.03	0.6	0.01	0.03	0.01	0.07	0.01	0.03	0.01
2	0	0.1	0.9	0	0	0	0	0	0	0
3	0	0.9	0	0.05	0.05	0	0	0	0	0
4	0	0.7	0	0.05	0.2	0.05	0	0	0	0
5	0	0.9	0	0	0.05	0.05	0	0	0	0
6	0	0.2	0	0	0	0	0.7	0.1	0	0
7	0	0.3	0	0	0	0	0.6	0.1	0	0
8	0	0.75	0	0	0	0	0	0	0.2	0.05
9	0	0.9	0	0	0	0	0	0	0.05	0.05

Table E.5: Occupant E.

Appendix F

Data Collection Experiment # 1: Symbols

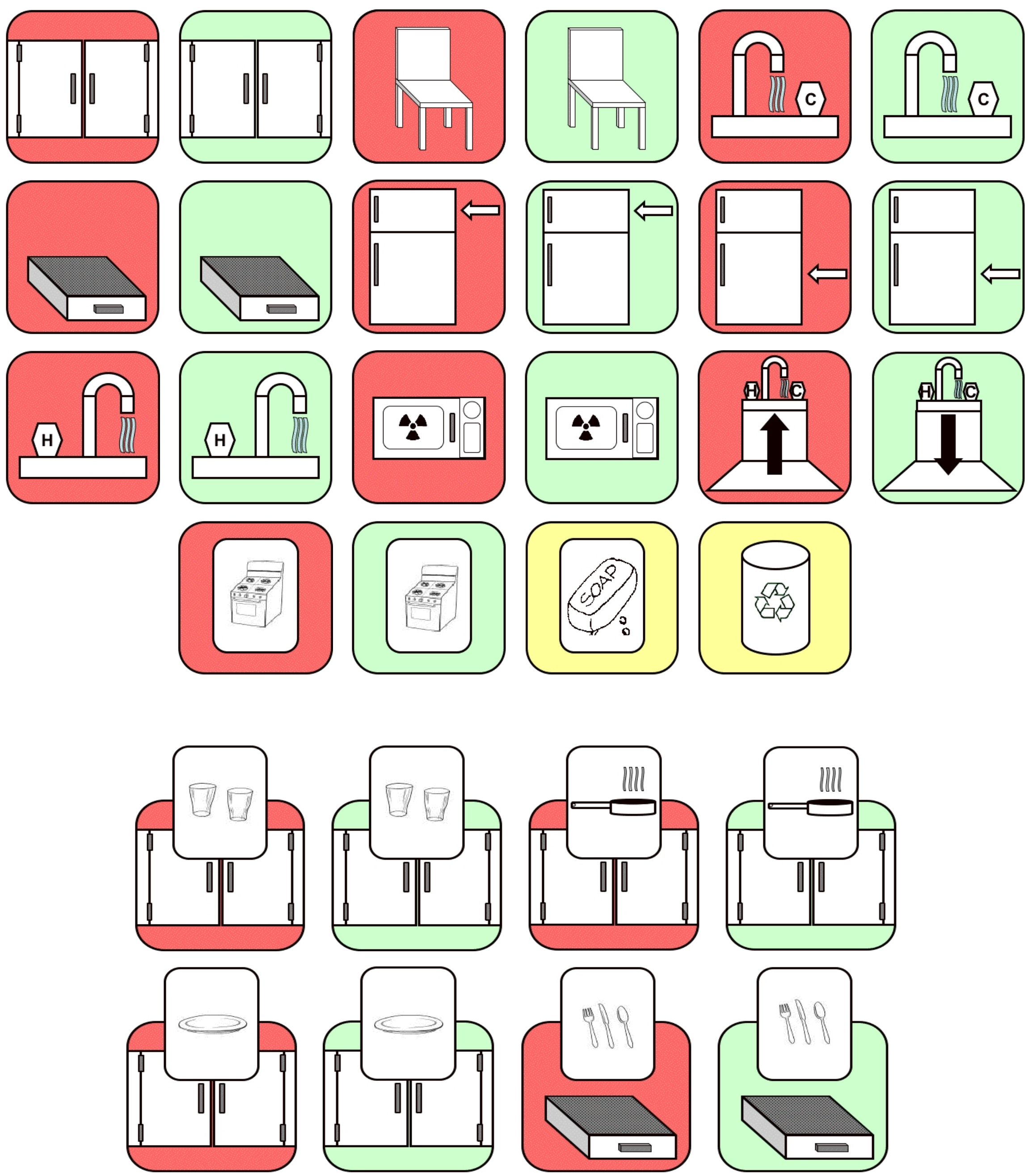

Figure F.1: Images used in CARS experiment # 1.

Appendix G

Data Collection Experiment # 2: Symbols

Figure G.1: Images used in CARS experiment # 2.

Appendix H

Data Collection Experiment # 2: Activity and Object Names

Object #	Object Name	Object #	Object Name
1	alarm system	35	lotion
2	bathroom door	36	magazine
3	bleach	37	microwave
4	bread	38	peanut butter
5	toilet brush	39	telephone
6	stove burner	40	phone book
7	cabinet	41	razor
8	calcium vitamin	42	TV remote
9	playing cards	43	shaving cream
10	cleaner	44	baby shirt
11	closet door	45	silverware
12	baby clothes	46	sink
13	clothes dryer	47	soap
14	coffee mug	48	softener
15	couch	49	spoon
16	kitchen counter	50	stereo
17	cutting board	51	drain stop
18	detergent	52	sugar
19	diaper	53	tea bag
20	dishwasher	54	teapot
21	bathroom fan	55	television
22	water faucet	56	thermostat
23	dental floss	57	tissue
24	baby formula	58	toilet
25	refrigerator	59	toilet paper
26	hair dryer	60	toothbrush
27	hair brush	61	toothpaste
28	high-chair	62	baby wipe
29	jelly	63	trash can
30	key	64	vacuum cleaner
31	butter knife	65	vacuum cleaner
32	laundry basket	66	vitamin c
33	light switch	67	washing machine
34	lipstick	68	water

Table H.1: The complete list of tagged objects from experiment # 2.

Activity #	Activity Name
1	Adjusting thermostat.
2	Apply make-up, lotion.
3	Brewing tea.
4	Brushing hair.
5	Brushing teeth.
6	Cleaning toilet.
7	Cleaning bathroom.
8	Cleaning kitchen.
9	Changing baby's diaper.
10	Dressing the baby.
11	Doing laundry.
12	Drinking water.
13	Making a PB&J sandwich.
14	Making a snack.
15	Playing solitaire.
16	Reading magazine.
17	Shaving face.
18	Taking vitamins.
19	Using dishwasher.
20	Using microwave.
21	Boil water with microwave.
22	Using telephone.
23	Using toilet.
24	Vacuuming.
25	Washing hands.
26	Watching TV.
27	None of the above.

Table H.2: The complete list of possible activities from experiment # 2.

Appendix I

Data Collection Experiment # 2: Instructions

INSTRUCTION SHEET

In this study you will play a short video game. In this game you will be shown objects that a person touched and then you will be asked to guess which activity they were performing. There will be 30 questions and the game should last about 10-15 minutes.

1. Watch until all images have scrolled onto the screen.
2. Select which activity you feel is most likely.
3. Select how confident you are about your choice.
4. Press the "SUBMIT" button to move on to the next question.

* You may pause the game at any time by pressing the "PAUSE" button.

* If the activity makes no sense or if it looks like two or more activities have appeared together, use the "NONE OF THE ABOVE" selection.

* People aren't perfect, and sometimes they might have accidentally touched unrelated objects in the middle of an activity. Try to keep this in mind when you label episodes.

* If it looks as though more than one activity fits, try to choose the most specific activity.

Thanks!!!

Figure I.1: Instruction sheet for experiment # 2.

Bibliography

[1] G. Abowd, C. Atkeson, A. Bobick, I. Essa, B. Mynatt, and T. Starner. Living laboratories: The Future Computing Environments group at the Georgia Institute of Technology. In *Proceedings of the 2000 Conference on Human Factors in Computing Systems (CHI)*, 2000.

[2] G. D. Abowd and E. D. Mynatt. Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction*, 7(1):29–58, 2000.

[3] M. Addlesee, R. Curwen, S. Hodges, J. Newman, P. Steggles, A. Ward, and A. Hopper. Implementing a sentient computing system. *IEEE Computer*, 34(8):50–56, August 2001.

[4] Administration on Aging. A profile of older Americans. http://www.aoa.gov/prof/Statistics/profile/2003/profiles2003.asp, 2003.

[5] L. V. Ahn and L. Dabbish. Labeling images with a computer game. In *Proceedings of CHI 2004*, pages 319–326, 2004.

[6] D. W. Albrecht, I. Zukerman, A. E. Nicholson, and A. Bud. Towards a Bayesian model for keyhole plan recognition in large domains. In *Proceedings of the Sixth International Conference on User Modeling (UM '97)*, pages 365–376, 1997.

[7] S. Allin, A. Barucha, J. Zimmerman, D. H. Wilson, M. Roberson, S. Stevens, H. Watclar, and C. Atkeson. Toward the automatic assessment of behavioral disturbances of dementia. In *UbiHealth 2003*, 2003.

[8] R. N. Anderson. Method for constructing complete annual US life tables, national center for health statistics. *Vital and Health Stat*, 2:129, 1999.

[9] P. Aoki, P. Syzmanski, J. Thornton, D. H. Wilson, and A. Woodruff. The mad hatter's cocktail party: A social mobile audio space that supports multiple simultaneous conversations. In *Human Factors in Computing Systems (CHI)*, 2003.

[10] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking. In *IEEE Transactions on Signal Processing*, 2002.

[11] Y. Bar-Shalom and T. Fortmann. Tracking and data association. *Mathematics in Science and Engineering*, 179, 1988.

[12] T. Barger, D. Brown, and M. Alwan. Health status monitoring through analysis of behavioral patterns. In *AI*IA 2003: Workshop on Ambient Intelligence*, 2003.

[13] L. Barrett and D. Barrett. An introduction to computerized experience sampling in psychology. *Social Science Computer Review*, 19(2):175–185, 2001.

[14] J. Beaudin, S. Intille, and E. M. Tapia. Lessons learned using ubiquitous sensors for data collection in real homes. In *Extended abstracts of CHI 2004*, pages 1359–1362, 2004.

[15] C. Beckman and S. Consolvo. Sensor configuration tool for end-users: Low-fidelity prototype evaluation # 1. Technical Report IRS-TR-03-009, Intel Research Seattle, July 2003.

[16] C. Beckmann, S. Consolvo, and A. LaMarca. Some assembly required: Supporting end-user sensor installation in domestic ubiquitous computing environments. In *Proceedings of UBICOMP 2004*, 2004.

[17] D. Beeferman, A. Berger, and J. Lafferty. Statistical models for text segmentation. *Machine Learning*, 34(1-3):177–210, 1999.

[18] M. Bennewitz, W. Burgard, and S. Thrun. Learning motion patterns of persons for mobile service robots. In *Proceedings of ICRA*, 2002.

[19] N. Bergman and A. Doucet. Markov chain Monte Carlo data association for target tracking. In *IEEE Conference on Acoustics, Speech and Signal Processing*, 2000.

[20] J. Bilmes. 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, 1997.

[21] J. Boger, J. Hoey, P. Poupart, C. Boutilier, G. Fernie, and A. Mihailidis. A planning system based on Markov decision processes to guide people with dementia through activities of daily living. In *IEEE Transactions on Information Technology in BioMedicine*, 2005.

[22] H. Bouma. Gerontechnology: A framework on technology and aging. In H. Bouma and J. Graafmans, editors, *Gerontechnology*, pages 1 – 6, Amsterdam, 1993. ISO Press.

[23] B. Brumitt, B. Meyers, J. Krumm, A. Kern, and S. Shafer. Easyliving: Technologies for intelligent environments. In *Handheld and Ubiquitious Computing*, 2000.

[24] L. Burgio, K. Scilley, J. M. Hardin, and C. Hsu. Temporal patterns of disruptive vocalization in elderly nursing home residents. *Int. Journal of Geriatric Psychiatry*, 16(1):378–386, 2001.

[25] L. Burgio, K. Scilley, J. M. Hardin, J. Janosky, P. Bonino, S. Slater, and R. Engberg. Studying disruptive vocalization and contextual factors in the nursing home using computer-assisted real-time observation. *Journal of Gerontology*, 49(5):230–239, 1994.

[26] C. Burkey. Environmental interfaces: Homelab. In *Proceedings of the Conference on Human Factors in Computer Systems*, 2000.

[27] Campo and Chan. Detecting abnormal behavior by real-time monitoring of patients. In *AAAI WS-02-02*, pages 8–12, 2002.

[28] Center for Aging Services and Technologies. Technology demo program. http://www.agingtech.org/techdemo.aspx, 2005.

[29] J. Chen, A. H. Kam, J. Zhang, N. Liu, and L. Shue. Bathroom activity monitoring based on sound. In *Proceedings of the International Conference on Pervasive Computing (PERVASIVE) 2005*, pages 47 – 61, 2005.

[30] M. C. Chenier. Review and analysis of caregiver burden and nursing home placement. *Geriatric Nursing*, 18(3):121–126, 1997.

[31] T. Choudhury and A. Pentland. Sensing and modeling human networks using the sociometer. In *Proceedings of the International Conference on Wearable Computing*, 2003.

[32] E. M. Clarke, O. Grumberg, and D. A. Peled. *Model Checking*. MIT Press, Boston, 2000.

[33] B. Clarkson, N. Sawhney, and A. Pentland. Auditory context awareness via wearable computing. In *Proceedings of the Perceptual User Interfaces Workshop*, 1998.

[34] D. Colbry, B. Peintner, and M. Pollack. Execution monitoring with quantitative temporal Bayesian networks. In *6th International Conference on AI Planning and Scheduling*, 2002.

[35] S. Consolvo, P. Roessler, and B. E. Shelton. The Carenet display: Lessons learned from an in home evaluation of an ambient display. In *Proceedings of the 6th Int'l Conference on Ubiquitous Computing: UbiComp '04*, pages 1–17, September 2004.

[36] S. Consolvo, P. Roessler, B. E. Shelton, A. LaMarca, B. Schilit, and S. Bly. Technology for care networks of elders. *IEEE Pervasive Computing Magazine: Successful Aging*, 3(2):22–29, April-June 2004.

[37] A. Culotta and A. McCallum. Confidence estimation for information extraction. In *Proc. of Human Lang. Tech. Conf. (HLT-NAACL)*, 2004.

[38] S. Czaja. Using technologies to aid the performance of home tasks. *Handbook of Human Factors and the Older Adult*, pages 311–334, 1997.

[39] L. Davis, K. Buckwalter, and L. Burgio. Measuring problem behaviors in dementia: Developing a methodological agenda. *Advances in Nursing Science*, 20(1):40–55, 1997.

[40] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag Telos, New York, 2001.

[41] T. V. Duong, H. H. Bui, D. Q. Phung, and S. Venkatesh. Activity recognition and abnormality detection with the switching hidden semi-Markov model. In *CVPR 2005*, 2005.

[42] K. P. Fishkin, H. Kautz, D. Patterson, M. Perkowitz, and M. Philipose. Guide: Towards understanding daily life via auto-identification and statistical analysis. In *UBIHEALTH 2003: The 2nd International Workshop on Ubiquitous Computing for Pervasive Healthcare Applications*, 2003.

[43] D. Fox, J. Hightower, H. Kautz, L. Liao, and D. Patterson. Bayesian techniques for location estimation. In *Proceedings of The 2003 Workshop on Location-Aware Computing*, October 2003. part of the 2003 Ubiquitous Computing Conference.

[44] D. Fox, J. Hightower, L. Liao, D. Schulz, and G. Borriello. Bayesian filtering for location estimation. *IEEE Pervasive Computing*, 2(3):24–33, July-September 2003.

[45] A. P. Glascock and D. M. Kutzik. Behavioral telemedicine: A new approach to the continuous non-obtrusive monitoring of activities of daily living. *Telemedicine Journal*, 6(10):33–45, 2000.

[46] V. Guralnik and K. Z. Haigh. Learning models of human behaviour with sequential patterns. In *AAAI Workshop 02*, pages 24 – 30, 2002.

[47] K. Z. Haigh and H. Yanco. Automation as caregiver: A survey of issues and technologies. In *AAAI-02 Workshop on Automation as Caregiver: The Role of Intelligent Technology in Elder Care*, pages 39 – 53, 2002.

[48] S. Helal, B. Winkler, C. Lee, Y. Kaddoura, L. Ran, C. Giraldo, S. Kuchibhotla, and W. Mann. Enabling location-aware pervasive computing applications for the elderly. In *First IEEE International Conference on Pervasive Computing and Communications*, page 531, 2003.

[49] J. Hightower and G. Borriello. Location systems for ubiquitous computing. *Computer*, 34(8):57–66, August 2001.

[50] L. Holmquist, J. Falk, and J. Wigstrm. Supporting group collaboration with interpersonal awareness devices. *Personal Technologies*, 3:13–21, 1999.

[51] T. Huang and S. Russell. Object identification in a Bayesian context. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97)*, 1997.

[52] C. Hue, J. L. Cadre, and P. Perez. The (MR)MTPF: Particle filters to track multiple targets using multiple receivers, August 2001.

[53] S. Intille and A. Bobick. A framework for recognizing multi-agent action from visual evidence. In *AAAI/IAAI*, pages 518–525, 1999.

[54] S. Intille and K. Larson. Designing and evaluating technology for independent aging in the home. In *Proceedings of the Int. Conference on Aging, Disability and Independence*, 2003.

[55] S. S. Intille. A new research challenge: Persuasive technology to motivate healthy aging. *Transactions on Information Technology in Biomedicine*, 8(3):235 – 237, 2004.

[56] S. S. Intille, K. Larson, J. S. Beaudin, J. Nawyn, E. M. Tapia, and P. Kashik. A living laboratory for the design and evaluation of ubiquitous computing technologies. In *2005 Conference on Human Factors in Computing Systems*, 2005.

[57] S. S. Intille, K. Larson, and E. M. Tapia. Designing and evaluating technology for independent aging in the home. In *Proceedings of the International Conference on Aging, Disability and Independence*, 2003.

[58] N. L. Johnson. *Continuous Multivariate Distributions*. John Wiley & Sons, 1972.

[59] M. Jug, J. Pers, B. Dezman, and S. Kovacic. Trajectory based assessment of coordinated human activity. In *ICVS 2003*, pages 534–543, 2003.

[60] R. L. Kahn, A. L. Goldfarb, M. Pollack, A. Peck, R. A. Kane, and R. L. Kane. Mental status questionnaire [MSQ]. In *Assessing the elderly: A practical guide to measurement*, pages 92 – 94, Lexington, Mass, 1981. Lexington Books.

[61] T. Kanade, R. Collins, A. Lipton, P. Burt, and L. Wixson. Advances in cooperative multi-sensor video surveillance. *Proceedings of DARPA Image Understanding Workshop*, 1:3–24, 1998.

[62] R. Karlsson and F. Gustafsson. Monte Carlo data association for multiple target tracking. In *IEEE Target Tracking: Algorithms and Applications*, October 2001.

[63] E. Kassner and B. Jackson. Determining comparable levels of functional disability. Technical Report Research Report IB32, American Association of Retired Persons, 1998.

[64] S. Katz, A. B. Ford, R. W. Moskowitz, B. A. Jackson, and M. W. Jaffe. Studies of illness in the aged: The index of ADL: A standardized measure of biological and psychosocial function. *Journal of the American Medical Association*, 185(12):914 – 919, 1963.

[65] E. Keogh, S. Chu, D. Hart, and M. Pazzani. *Data Mining in Time Series Databases*, chapter Segmenting Time Series: A Survey and Novel Approach. World Scientific Publishing Company, 2003.

[66] Z. Khan, T. Balch, and F. Dellaert. Efficient particle filter-based tracking of multiple interacting targets using an mrf-based motion model. In *Proceedings of ICRA*, 2003.

[67] I. Korhonen, J. Parkka, and M. van Gils. Health monitoring in the home of the future: Infrastructure and usage models for wearable sensors. *IEEE Eng Med Biol 2003*, 22(3):66 – 73, 2003.

[68] M. P. Lawton. Aging and performance of home tasks. *Human Factors*, 32:527 – 536, 1990.

[69] S. Lee and K. Mase. Activity and location recognition using wearable sensors. In *First IEEE International Conference on Pervasive Computing and Communications*, pages 24–32, 2002.

[70] R. A. Levine and G. Casella. Implementations of the Monte Carlo EM algorithm. *Journal of Computational and Graphical Statistics*, 2000.

[71] L. Liao, D. Fox, and H. Kautz. Location-based activity recognition using relational markov networks. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2005.

[72] Living Independently. QuietCare system (TM). www.livingindependently.com, 2005.

[73] C. B. Mayhorn, W. A. Rogers, and A. D. Fisk. Designing technology based on cognitive aging principles. In S. Kwon and D. Burdick, editors, *Gerontechnology: Research and Practice in Technology and Aging*, New York, 2004. Springer.

[74] S. Melenhorst, A. Fisk, E. Mynatt, and W. Rogers. Potential intrusiveness of aware home technology: Perceptions of older adults. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, 2004.

[75] J. P. Mellor and C. Hebb. Aging in place position paper. In *Extended Abstracts of CHI 2004: Workshops Home Technologies to Keep Elders Connected*, 2004.

[76] D. Meyer. Human gait classification based on hidden Markov models. In *3D Image Analysis and Synthesis*, pages 139–146, 1997.

[77] A. Mihailidis and G. Fernie. Context-aware assistive devices for older adults with dementia. *Gerontechnology*, 2(2):173 – 189, 2003.

[78] A. Mihailidis, G. Fernie, and W. Cleghorn. The development of a computerized cueing device to help people with dementia to be more independent. *Technology and Disability*, 13:23 – 40, 2000.

[79] D. Moore, I. Essa, and M. Hayes. Exploiting human actions and object context for recognition tasks. In *Proceedings of IEEE International Conference on Computer Vision 1999 (ICCV99)*, 1999.

[80] M. Mozer. The neural network house: An environment that adapts to its inhabitants. In *Proceedings of the American Association for Artificial Intelligence Spring Symposium on Intelligent Environments*, pages 110–114, 1998.

[81] E. D. Mynatt, I. Essa, and W. A. Rogers. Increasing the opportunities for aging in place. In *Proceedings of CUU 2000*, pages 65 – 71, 2000.

[82] E. D. Mynatt and W. A. Rogers. Developing technology to support the functional independence of older adults. *Aging International*, 2002.

[83] E. D. Mynatt, J. Rowan, A. Jacobs, and S. Craighill. Digital family portraits: Supporting peace of mind for extended family members. In *Proceedings of the Conference on Human Factors in Computing Systems: CHI 01*, pages 333 – 340, April 2001.

[84] M. Ogawa, R. Suzuki, S. Otake, T. Izutsu, T. Iwaya, and T. Togawa. Long-term remote behavioral monitoring of the elderly using sensors installed in domestic houses. In *Engineering in Medicine and Biology. 24th Annual Conference and*

the Annual Fall Meeting of the Biomedical Engineering Society, volume 3, pages 23–26, 2002.

[85] M. Ogawa and T. Togawa. The concept of the home health monitoring. In *Proceedings of Enterprise Networking and Computing in Healthcare Industry. Healthcom 2003*, pages 71–73, 2003.

[86] T. Ogawa, S. Yoshino, M. Shimizu, and H. Suda. A new in-door location detection method adopting learning algorithms. In *PerCom 2003*, pages 525–530, 2003.

[87] N. Oliver, E. Horvitz, and A. Garg. Layered representations for human activity recognition. In *Fourth IEEE International Conference on Multimodal Interfaces*, pages 3–8, 2002.

[88] M. Ostendorf, V. V. Digalakis, and O. A. Kimball. From HMMs to segment models: A unified view of stochastic modeling for speech recognition. *IEEE Transactions on the Speech Audio Process*, 4(5):360 – 378, 1996.

[89] H. Pasula, S. Russell, M. Ostland, and Y. Ritov. Tracking many objects with many sensors. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI)*, 1999.

[90] D. Patterson, D. Fox, H. Kautz, and M. Philipose. Expressive, tractable and scalable techniques for modeling activities of daily living. In *UbiHealth 2003: The 2nd International Workshop on Ubiquitous Computing for Pervasive Healthcare Applications*, 2003.

[91] D. Patterson, L. Liao, D. Fox, and H. Kautz. Inferring high-level behavior from low-level sensors. In *Proceedings of UBICOMP 2003*, October 2003.

[92] M. Philipose, K. Fishkin, M. Perkowitz, D. Patterson, H. Kautz, and D. Hahnel. Inferring activities from interactions with objects. *IEEE Pervasive Computing Magazine: Mobile & Ubiquitous Systems*, 3(4):50–57, October-December 2004.

[93] M. Philipose, K. Fishkin, M. Perkowitz, D. Patterson, H. Kautz, and D. Hahnel. Inferring activities from interactions with objects. *IEEE Pervasive Computing Magazine*, 3(4):50–57, 2004.

[94] J. Pineau, M. Montemerlo, N. Roy, S. Thrun, and M. Pollack. Towards robotic assistants in nursing homes: Challenges and results. *Robotics and Autonomous Systems*, 2003.

[95] M. K. Pitt and N. Shephard. *Sequential Monte Carlo Methods in Practice*, chapter Auxiliary Variable Based Particle Filters. Cambridge University Press, 2000.

[96] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[97] B. Rhodes. The wearable remembrance agent: A system for augmented memory. In *Proceedings of the 1st International Symposium on Wearable Computers*, pages 123–128, 1997.

[98] Robert Wood Johnson Foundation. Chronic care in America: A 21st century challenge. www.rwjf.org/files/publications/ other/ChronicCareinAmerica.pdf, 1996.

[99] W. A. Rogers, B. Meyer, N. Walker, and A. D. Fisk. Functional limitations to daily living tasks in the aged: A focus group analysis. *Human Factors*, 40:111 – 125, 1998.

[100] J. Rondoni. Context-aware experience sampling for the design and study of ubiquitous technologies. Master's thesis, Massachusetts Institute of Technology, September 2003.

[101] J. Rowan and E. D. Mynatt. Digital family portrait field trial: Support for aging in place. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: CHI 05*, pages 521 – 530, April 2005.

[102] C. Salzman. Medication compliance in the elderly. *Journal of Clinical Psychiatry*, 56(18 - 22), 1995.

[103] I. D. Shlomo Argamon-Engelson. Committee-based sample selection for probabilistic classifiers. In *Journal of Artificial Intelligence Research*, volume 11, pages 335–360, 1999.

[104] H. Sidenbladh and M. Black. Learning image statistics for Bayesian tracking. *IEEE International Conference on Computer Vision (ICCV)*, 2:709–716, 2001.

[105] T. Starner. Visual recognition of American sign language using hidden Markov models. Master's thesis, MIT Media Lab, 1994.

[106] K. Sycara, K. Decker, A. Pannu, M. Williamson, and D. Zeng. Distributed intelligent agents. *IEEE Expert*, 11(6):36–46, December 1996.

[107] P. Tang and T. Venables. Smart homes and telecare for independent living. *Journal of Telemedicine and Telecare*, 6(1):8 – 14, 2000.

[108] E. M. Tapia, S. S. Intille, and K. Larson. Activity recognition in the home setting using simple and ubiquitous sensors. In *Proceedings of PERVASIVE 2004*, volume LNCS 3001, pages 158–175, 2004.

[109] S. Thrun and J. Langford. Monte Carlo hidden Markov models. Technical Report TR CMU-CS-98-179, Carnegie Mellon University, 1997.

[110] U.K. National Office of Statistics. Focus on older people. http://www.statistics.gov.uk/focuson/olderpeople/, 2004.

[111] U.S. Census Bureau. Population projections of the United States by age, sex, and Hispanic origin 1995 to 2050. http://www.census.gov, 2005.

[112] K. VanHaitsma, P. Lawton, M. Kleban, J. Klapper, and J. Corn. Methodological aspects of the study of streams of behavior in elders with dementing illness. *Alzheimer Disease and Associated Disorders*, 11(4):228–238, 1997.

[113] D. Wan. Magic medicine cabinet: A situated portal for consumer healthcare. In *First International Symposium on Handheld and Ubiquitous Computing (HUC '99)*, pages 27 – 29, 1999.

[114] J. D. Weeks. National center for health statistics and national institute on aging: The second longitudinal study of aging. http://www.cdc.gov/nchs/lsoa.htm, 2005.

[115] G. Wei and M. Tanner. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, 85(411):699–704, 1990.

[116] D. H. Wilson and C. Atkeson. The narrator : A daily activity summarizer using simple sensors in an instrumented environment. In *The Fifth International Conference on Ubiquitous Computing (UBICOMP) 2003 Demonstrations, Seattle, USA*, 2003.

[117] D. H. Wilson and C. Atkeson. Simultaneous tracking and activity recognition using many anonymous, binary sensors. In *Proceedings of the International Conference on Pervasive Computing (PERVASIVE) 2005*, pages 62–79, 2005.

[118] D. H. Wilson, S. Consolvo, K. P. Fishkin, and M. Philipose. Current practices for in-home monitoring of elders' activities of daily living: A study of case managers. Technical Report IRS-TR-05-015, Intel Research Seattle, September 2005.

[119] D. H. Wilson and A. C. Long. A context-aware recognition survey for data collection using ubiquitous sensors in the home. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI) 2005: Late Breaking Results*, pages 1865–1868, 2005.

[120] D. H. Wilson and M. Philipose. Maximum a posteriori path estimation with input trace perturbation: Algorithms and application to credible rating of human activities. In *Proceedings of IJCAI 2005*, 2005.

[121] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of human body. *Machine Intelligence*, 19(1):780–785, 1997.