POMDP Models for Assistive Technology

Jesse Hoey
Dept. of Computer Science
University of Toronto
6 King's College Rd
Toronto, Canada, M5S3H5

Pascal Poupart
Sch. of Computer Science
University of Waterloo
200 University Avenue W.
Waterloo, Canada, N2L3G1

Craig Boutilier
Dept. of Computer Science
University of Toronto
6 King's College Rd
Toronto, Canada, M5S3H5

Alex Mihailidis
Dept of Occupational Therapy
University of Toronto
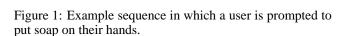500 University Ave.
Toronto, Canada M5G1V7

Abstract

This paper presents a general decision theoretic model of interactions between users and cognitive assistive technologies for various tasks of importance to the elderly population. The model is a partially observable Markov decision process (POMDP) whose goal is to work in conjunction with a user towards the completion of a given activity or task. This requires the model to monitor and assist the user, to maintain indicators of overall user health, and to adapt to changes. The key strengths of the POMDP model are that it is able to deal with uncertainty, it is easy to specify, it can be applied to different tasks with little modification, and it is able to learn and adapt to changing tasks and situations. This paper describes the model, gives a general learning method which enables the model to be learned from partially labeled data, and shows how the model can be applied within our research program on technologies for wellness. In particular, we show how the model is used in three tasks: assisted handwashing, health and safety monitoring, and wheelchair mobility. The paper gives an overview of ongoing work into each of these areas, and discusses future directions.

Introduction

A growing area of activity in health technology is support systems for older adults, possibly with cognitive or physical disabilities, who want to continue to live independently in their own homes i.e. age-in-place. Such systems are typically engineered for a certain task to provide guidance, assistance, or emergency response (Mihailidis & Fernie 2002; LoPresti, Mihailidis, & Kirsch 2004). However, this approach is labour intensive, and the resulting systems tend to have no capacity to adapt over time or to different users or tasks. In this paper, we discuss an approach to this problem: a ubiquitous modeling technique that can adapt to users over time. The idea is to have a single model and learning technique that can be easily applied to different tasks, without the need to re-engineer the model.

A typical task requiring assistance consists of four principal elements. We discuss these elements here in the context of the handwashing task for cognitively disabled people, who typically require assistance from a human caregiver to wash their hands. An example of this is shown in Figure 1, which shows key frames from about 15 seconds of a video

"I want you to use some soap now"

Figure 1: Example sequence in which a user is prompted to put soap on their hands.

of a user washing their hands, assisted by a human caregiver. First, the *task state* is a characterisation of the high-level state of the user, and is related to the goals in the task. For example, handwashing can be described by task states that describe whether the hands are wet or dry, dirty, soapy or clean. In Figure 1, the user's hands are dirty and wet at frame 1331, but become soapy and wet by frame 1745. Second, the *behavior* of the user is the course of action the user takes to change the task state. Common behaviors during handwashing may be things like *rinsing hands* or *using soap*, as in Figure 1. Third, the *caregiver's action* is what the caregiver does to help the user through the task. During handwashing, these actions are typically verbal prompts or reminders such as the *"I want you to use some soap now"* in Figure 1. However, as we will show, actions can include more general dialogue iterations, calls to other response systems, or physical control of related systems. The fourth element, the users *attitude*, is the cognitive state of the user, such as their level of responsiveness, attention, frustration with the system, and overall level of health. The user's expeditious reaction to the prompt in Figure 1, for example, might give us an indication that they are responsive, and are attending to the prompting system. Over a longer time period, the user's overall level of health may change. For example, their medical condition might take a turn for the worse, requiring attention from a professional. Such a change may be noticeable in their responses and behaviors.

Our goal is then to design a model of the interactions between these four elements, and to optimize an automated caregiving strategy by maximising (over the actions) some

notion of utility over the possible outcomes. The model must be able to deal with uncertainty in the effects of actions and in sensor measurements, it must be able to tailor to specific individuals and circumstances, it must be able to trade off various objective criteria (e.g., task completion, caregiver burden, user frustration and independence), and it must be relatively easy to specify. A partially observable Markov decision process (POMDP), a decision theoretic model which has recently received much attention in the AI community, fulfills these constraints.

The general POMDP we present models the *task* and the *attitude* as consequences of the *behavior* of the user, which is a reaction to the *actions* of the caregiver. We claim that the *task* will tend to be simple to specify, and can be defined by a non-specialised person, while the *attitude* will require expert knowledge, but will tend to generalise across tasks. On the other hand, the *behaviors* will be much more difficult to specify. The standard approach to specifying models of behaviors occurs in two phases. In the first phase, expert knowledge is used to define the behaviors that will occur, and a supervised classifier is trained to recognise these statically defined behaviors. In the second phase, a model of the relationship between the behaviors (as given by the learned classifier) and the task is learned. The problem with this approach is the need to specify behaviors *a priori*. This is a labor-intensive task requiring expert knowledge. Further, different individuals perform the same behaviors in different ways, exhibit different behaviors to perform the same task, and change their behaviors over time, usually as a result of their changing state of health. These considerations make it very difficult in many cases to define a single set of recognisable behaviors. This aspect is emphasised in the rehabilitation literature extensively (LoPresti, Mihailidis, & Kirsch 2004). Finally, users require assistance for very different aspects of a task, and so recognising all possible behaviors may be very wasteful. Our approach is to *discover* the behaviors that are being exhibited, and to *learn* their relationship to the task simultaneously (Hoey & Little 2004). This learning method has the dual advantage of not requiring extensive *a priori* knowledge and training, and of being capable of adapting to different users in different situations.

The paper first describes related work, then describes a general POMDP model, including the observation function and the methods for specifying, learning, and solving the POMDP. We will then discuss the application of this general model in our own research program on technologies for wellness, where we focus on the development of systems that can assist older adults in a variety of contexts – specifically, completion of activities of daily living (ADL), health and safety monitoring, and improved mobility.

Related Work

Cognitive Assistive technologies have been the subject of much research outside the artificial intelligence community, and are reviewed in (LoPresti, Mihailidis, & Kirsch 2004). Most relevant to our work, a system for monitoring handwashing using a ceiling-mounted camera was demonstrated in (Mihailidis, Barbenel, & Fernie 2004). The user was required to wear a patterned bracelet, the location of which

was determined by a pattern recognition algorithm. The resulting location was then input to a neural network for the recognition of predefined behaviors. This system was invasive and was not learned from data. A POMDP model for a scheduling system for the care of the elderly was described in (Rudary, Singh, & Pollack 2004). Their work is part of the *Autominder* project, which also includes design of a robotic nursing assistant that uses a POMDP for navigation purposes (Montemerlo *et al.* 2002). POMDPs have recently been applied to dialogue management in care facilities (Roy, Gordon, & Thrun 2003).

There has been significant progress in learning patterns of activity from a person's positional data. These include data mining techniques for discovering sequences of activities from discrete data (Guralnik & Haigh 2002), and learning the parameters of a hierarchical hidden Markov model to explain GPS data of outdoor transportation patterns (Liao *et al.* 2004). We are learning a similar model, but explicitly add system actions and model video sequences directly instead of only positional data. Other researchers use supervised techniques to build models of meeting dynamics (Rybski & Veloso 2004), office activity (Nguyen *et al.* 2003), and other in-home activities (Hamid, Huang, & Essa 2003).

Our previous work, (Hoey & Little 2004; Hoey *et al.* 2005), showed how to learn the parameters of a fully observable Markov decision process, while *discovering* models of behaviors. We have also applied some preliminary versions of the systems presented here to handwashing (Boger *et al.* 2005b), emergency response (Mihailidis *et al.* 2005) and fall detection (Lee & Mihailidis 2005).

General Model

A discrete-time POMDP consists of: a finite set S of states; a finite set A of actions; a stochastic transition model $\Pr : S \times A \rightarrow \Delta(S)$, with $\Pr(t|s, a)$ denoting the probability of moving from state s to t when action a is taken; a finite observation set O; a stochastic observation model with $\Pr(o|s)$ denoting the probability of making observation o while the system is in state s; and a reward assigning reward $R(s, a, t)$ to state transition s to t induced by action a. Figure 2(a) shows the POMDP as a Bayesian network. Given a specific POMDP, our goal is to find a *policy* that maximizes the expected discounted sum of rewards attained by the system. Since the system state is not known with certainty, a policy maps either *belief states* (i.e., distributions over S) or action-observation *histories* into choices of actions. We will not delve into details of POMDP solution methods, but note that current research has enabled the approximate solution of very large POMDPs (Poupart & Boutilier 2004; Hoey & Poupart 2005), and we are confident that our current work will allow us to solve the POMDPs we are presenting in this paper. We refer to (Lovejoy 1991) for an overview of POMDP concepts and algorithms.

Figure 2(b) shows the same model, except that the state, S, has been factored into three sets of variables: *task* (S_h), *attitude* (S_p) and *behavior* (S_b). Here we describe each of these sets, as well as the actions of the system, A, and the observations from which the state of the model is inferred.
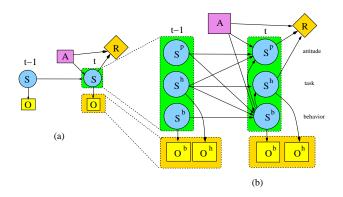
Figure 2: (a) Two time slices of general POMDP. The state, S, is modified by action A, and produces observation O. (b) Two slices of factored POMDP for modeling user behavior during interaction with cognitive assistive technology.

Task, S_b: A characterisation of the domain in which the task is taking place in terms of a set of high-level variables. These will typically describe the physical state of items in the task, such as the cleanliness of hands during handwashing, or the location of a user in a mobility aid. These variables need to be specified for each task, but they characterise the world at a sufficiently high level to make this accessible to a wide variety of non-technical people.

Attitude, S_p: variables describe the cognitive state of the user. This may include the level of dementia and the current responsiveness, or perhaps the level of frustration the user is experiencing with the system itself. The *attitude* variables essentially describe internal properties of the user that generalise across tasks, and would be specified by experts for the general system, and then carried over and possibly adapted for each task through learning.

Behavior, S_b: The task states and the user's attitude are changed by the user's *behavior*, S_b. The user's *behavior* evolves depending on the previous states of the *attitude* and the *task* as well as the system's action, A. The *behaviors* are probably the most difficult to manually specify, and so are learned by the system from data.

Action, A: The actions of system modify the behavior, attitude, and the task. These actions could be verbal prompts, calls to human caregivers or other response systems, or physical changes to the environment.

Observations, O: *Task* and *behavior* variables generate observations, O_h and O_b, respectively. We will assume here that these observations are generated from non-invasive sensors, such as cameras (in which case the observations are video streams), microphones (audio streams), or environmental switches such as thermostats.

Jointly, $S = \{S_b, S_p, S_h\}$ is known as the state. The transition function here, $Pr(S_t|S_{t-1}, A_t) = Pr(S_{bt}, S_{pt}, S_{ht}|S_{bt-1}, S_{pt-1}, S_{ht-1}, A_t)$, is written as a product of three terms, as follows. $Pr(S_{bt}|S_{t-1}, A_t)$ gives the expected behavior of the user given the previous state and the system action. $Pr(S_{pt}|S_{bt}, S_{pt-1}, S_{ht-1}, A_t)$ gives

the expected user state given the current behavior, the previous attitude and task states and the system action. $Pr(S_{ht}|S_{ht-1}S_{bt})$ gives the expected task state given the current behavior and the previous task state. Notice that the only conditional independencies introduced here are in this last distribution: the task state is independent of the attitude, S_p, and the action, A. The idea is that changes in the *task* states are caused by the *behaviors* of the user, independently of the user's attitude or the system's actions. The action, A, only affects the *behavior* and *attitude* of the user, which in turn may cause changes to the *task*.

The observations $O = \{O_b, O_h\}$ are generated by the *task* and *behavior* variables, S_h and S_b, respectively, through some observation functions $Pr(O_h|S_h)$ and $Pr(O_b|S_b)$. These distributions can be of many different types, depending on what the observations are. In general, however, the time scales at which observations occur will be of much shorter duration than those at which task or attitude change. Observations of behaviors will typically be frames from some video camera (at 30Hz), or some segments of an audio stream (at 10kHz), whereas the task states will only change every few seconds. For example, during handwashing, a typical behavior may be "putting soap on hands", which may take a few seconds to perform, and result in 30 video frame observations (e.g. Figure 1), but only cause a single change of the *task* state: the hands become "soapy". Thus, the observation functions may introduce some hierarchical structure into the model.

Specifying, Learning, and Solving POMDPs

There are three steps to obtaining a working POMDP-based system: specification, learning and solving. The POMDP is first *specified* by defining a set of *task* and *attitude* variables, the observation space (what the sensors will be reporting to the system), a set of actions the system can take, a reward function, and the conditional dependencies in the associated Bayesian network (Figure 2). The transition function $Pr(S_{t+1}|S_t, A)$ and the observation function $Pr(O|S)$ are then *learned* from data, and the resulting POMDP is then *solved* to yield a policy of action, mapping observations to actions. Ideally, these last two steps are combined in one, and the system learns both the model and the policy online as it interacts with users, known as reinforcement learning (RL). However, online learning requires a great deal of *exploration* for the system to discover how the domain works, so that it can exploit its knowledge (and reach goals, etc.). This exploration can be very costly, especially in a clinical setting with cognitively or physically disabled users. *Bayesian* reinforcement learning optimises the tradeoff between exploration and exploitation by including the model parameters as part of the state that is being estimated during online inference (Dearden, Friedman, & Russell 1998). The intuition is that the system should include its learning task as one of its goals (along with the goals defined by the domain itself). In a principled way, Bayesian RL also allows incorporation of prior knowledge to reduce the need for exploration. We integrate this prior knowledge in three stages. First, we elicit a reward function (either manually or automatically) from users, experts and caregiving profes-

sionals. Second, we elicit a partial transition function that describes the dynamics of the *attitude* variables from data. This information will be gathered from experts (e.g. psychologists) who know how the cognitive states of a target population change in situations of interest. Third, we learn a prior model offline from a set of training data gathered using an existing automated system or a human caregiver (discussed below). Finally, we solve the model using an offline dynamic programming technique (Hoey & Poupart 2005), to yield an initial policy of action. Bayesian reinforcement learning can then be used to adjust this model for each task and each user. In the following, we only describe the offline learning of the prior model, which is designed to yield a model sufficient for deployment in a clinical setting. We do not discuss the subsequent application of Bayesian RL further in this paper.

Our method first gathers a corpus of training data using an existing automated system, or using a human caregiver. The training data is then be partially labeled for the *task* variables by non-experts. Finally, the model is learned using the expectation-maximization (EM) algorithm (Bengio & Frasconi 1996), which finds the model parameters that maximize the posterior density of all observations and the model. Denote the set of observations in the training set $\mathbf{O}$, the caregiver's prompts $\mathbf{A}$, and task states $\mathbf{S_h}$. The learning problem is to find the set of parameters $\mathbf{\Theta}^*$ that maximizes $Pr(\mathbf{O}, \mathbf{S_h}, \mathbf{A}, \mathbf{\Theta})$, subject to constraints on the parameters, which involves integrating over all possible user behaviors $\mathbf{S_b}$ and attitudes $\mathbf{S_p}$ The EM algorithm eases this maximization by writing it

$$\arg\max_{\mathbf{\Theta}} \left[\sum_{\mathbf{S_b},\mathbf{S_p}} Pr(\mathbf{S_b}\mathbf{S_p}|\mathbf{OS_h}\mathbf{\Theta}') \log Pr(\mathbf{OS}|\mathbf{\Theta}) \right.$$
$$\left. + \log Pr(\mathbf{\Theta}) \right]$$

The "E" step of the EM algorithm is to compute the expectation over the hidden behaviors and attitude, $Pr(\mathbf{S_b}\mathbf{S_p}|\mathbf{OS_h}\mathbf{\Theta}')$, given $\mathbf{\Theta}'$, a current guess of the parameter values. The "M" step is then to perform the maximization which, in this case, can be computed analytically by taking derivatives with respect to each parameter, setting them to zero and solving for the parameters (Bengio & Frasconi 1996). One of the most interesting aspects of this learning method is that it *discovers* classes of behaviors that occur in the data (Hoey & Little 2004).

Activity Models

We will demonstrate how to apply the POMDP model to three tasks: prompting assistance during handwashing for people with Alzheimer's, emergency response for the elderly, and wheelchair control for the cognitively disabled. For each task, we introduce the domain, and then give an overview of the *task* and *attitude* variables, and of the system actions. We then describe the observations and observation functions, the transition dynamics, and the reward function. Finally, we describe some of our current work.

Handwashing

Older adults living with cognitive disabilities (such as Alzheimer's disease or other forms of dementia) have difficulty completing activities of daily living (ADLs), and are usually assisted by a human caregiver who prompts them when necessary. Assistive technology will allow this elderly population to age-in-place by non-invasively monitoring users in their homes during ADL, providing guidance or assistance when necessary (LoPresti, Mihailidis, & Kirsch 2004). In the handwashing ADL, the user needs to get his hands clean by progressing through stages that include using soap, turning the water on and off, rinsing and drying his hands. A caregiver monitors the progress of the user, issuing reminders or prompts at appropriate times.

Task The handwashing task can be described by three primary task variables: *hands_clean*, which can be {*dirty, soapy, clean*}, *hands_wet*, which can be {*wet, dry*}, and *water_flow*, which can be {*on, off*}. We assume the hands start *dirty* and *dry*, and the goal is to get them *clean* and *dry*, which can only happen if they become *soapy* and *wet* at some intermediate time. The water starts *off* and must be *off* for task completion. Only the water flow has a corresponding observation (e.g. a water flow switch) from which it can be observed. Other task variables are involved with timing issues (such as how much time the system has waited for, how many prompts have been issued with no response, etc), and with issues related to the user's current performance (such as how many times they have regressed in the task). See (Boger *et al.* 2005b) for details.

Attitude The attitude of the user can play a critical role in their ability to respond to prompts. We use two variables for handwashing: *attentiveness*, describing how likely they are to attend to the an audio prompt, and *responsiveness*, describing how likely they are to respond to a prompt if they hear it.

Actions The system actions are the possible reminders that can be given to the user in the form of audible cues, corresponding to the canonical steps of handwashing: *turn on water*, *wet hands*, *use soap*, *dry hands* and *turn off water*. The prompts can be issued at three levels of specificity. There is also one null action where the system waits (to give more independence to the user), and one action of calling for human assistance. This last action is important for such a system: it must know when its limitations have been reached. The goal is to reduce burden on caregivers, not to replace caregivers entirely. Audio cues have been found to be sufficient to decrease caregiver burden in previous studies (Mihailidis, Barbenel, & Fernie 2004).

Observations The user's behaviors are inferred from videos taken from an overhead camera. These behaviors have a temporal extent over multiple video frames and require a hierarchical observation function. The behaviors are learned from data, but typically correspond to the usual activities in handwashing, such as using the soap, or reaching for the towel (Hoey *et al.* 2005).

Dynamics The transition dynamics encode the progression

of the handwashing task. For example, hands become clean if they are soapy and the user performs a behavior of rinsing. The probability of an appropriate user behavior as a response to prompts increases with prompt specificity and with user responsiveness.

Reward The reward function gives a positive reward for task completion, small negative rewards for each prompt (encoding the preference of users to complete tasks on their own), and a large negative reward for calling a human caregiver.

We have experimented with two simplified versions of this POMDP. The first version uses a simplified set of variables (ignoring attitude and timing issues), and shows how models of behaviors could be learned from data, and how they could be subsequently used to monitor user progress in the task (Hoey *et al.* 2005). The second version of this POMDP uses an *a-priori* specification of the model parameters and a simplified observation function (Boger *et al.* 2005a; 2005b). Observations of hand positions are derived from video frames using skin color analysis (Mihailidis, Carmichael, & Boger 2004), and behaviors are defined as the most likely location of the hands over each temporal sequence. Despite the size and complexity of the model (over 50 million states), we were able to solve the POMDP, producing (approximately) sequentially optimal policies for ADL prompting. Clinical trials have been completed with a fully observable version of this model. More trials are planned to validate the full POMDP. Figure 3 shows an example in which a person with dementia is prompted by the fully observable version of the POMDP model during a clinical trial.

Health Monitoring and Emergency Response

Aging-in-place is difficult for adults who live alone, possibly in rural areas where assistance may not be readily available (Mynatt, Essa, & Rogers 2000; Mihailidis & Fernie 2002). For example, they may have a heart attack or stroke, fall, or become somehow incapacitated without the ability to call for help. In addition to these spontaneous adverse events, there may also be a gradual decline in overall health status, both physical and cognitive, that further places older adults at risk. There are currently several attempts to address these issues and support aging-in-place through the use of technological solutions, the most common being emergency response systems and health monitoring devices. The most common emergency response system (ERS) is the telephone-based personal emergency response system (PERS), which consists of the subscriber wearing a small help button as a necklace or wristband, and a home communicator that is connected to a residential phone line. In the event of an emergency, the subscriber presses the help button and is connected to a live emergency response centre, which arranges for appropriate help, such as calling paramedics or the person's family. Remote health monitoring devices have also been developed that measure and track various physiological parameters, such as pulse, skin temperature, and blood pressures (Asada *et al.* 2003).

However, these systems are invasive: they require the user to wear the device at all times and/or to manually take the required measurements. Many of the PERS and physiological-based monitoring systems are inappropriate, obtrusive and difficult for an older adult to operate for various reasons (e.g. require effort from the user, long training periods, etc.), and they become ineffective during more serious emergency situations (e.g. the person has a stroke). As a result, new systems are being developed that do not require manual interaction from the user, and that use non-physiological measures to determine health (Lee & Mihailidis 2005). Here we show how the POMDP model we have discussed can be used to model the health monitoring and emergency response tasks.

The concept is to have cameras, microphones and speakers installed throughout a home, monitoring the user's activities during the day (for health monitoring) and providing assistance or emergency response when needed. In this case, there are two inputs: video and audio, which relate to visually observable behaviors of the user and their speech acts, respectively. This POMDP combines elements of human activity modeling (Wren *et al.* 1997; Oliver, Horvitz, & Garg 2002) and dialogue management (Roy, Gordon, & Thrun 2003; Williams, Poupart, & Young 2005).

Task Task variables include descriptors of the person's location, the time of day, the frequency of visits to different locations, and whether they have fallen or not.

Attitude Attitude variables here can include overall level of health (such as fatigue, recurrence of disease symptoms, etc), responsiveness to prompts, current alertness level.

Actions Possible actions are questions or speech acts (dialogue), or calls to neighbors, caregivers, or paramedics.

Observation There are two *behavior* variables: the visually observable actions, such as motion (from room to room or within a room) and unusual events (falls), and the audible actions, such as speech acts or other noises. The visual and audio behaviors of a user are related to the observations of video (from cameras) and audio (from microphones) through separate observation functions. Video observations are processed using adaptive background subtraction and blob analysis (Wren *et al.* 1997). Audio observations are processed using the Sphinx open-source speech recognition engine (Walker *et al.* 2004), which gives the ability to assess the observation probability $Pr(o|s)$, where o is the audio stream and s are the speech tokens, which correspond to the behaviors, S_b in our model. Since the speech recognition engine uses hidden Markov models, this will implicitly be a hierarchical observation function. The information from the visual modality has also been used to help the speech recognition using a microphone array and a beam-forming algorithm. This allows sources of noise to be attenuated, so that the audio signal comes mainly from the area in which a person was observed to have fallen.

Dynamics The model encodes two distinct, but related, functions. First, the monitoring of the overall level of health of the person is done by the dynamics of the *attitude* and *task* variables. For example, if the frequency of visits to a particular location (e.g. the bathroom) sud-

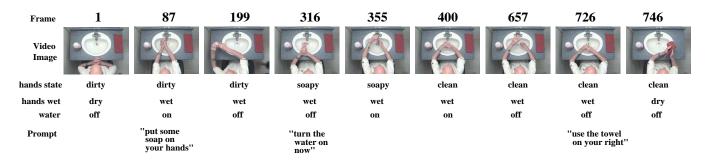| Frame	1	87	199	316	355	400	657	726	746
Video Image									
hands state	dirty	dirty	dirty	soapy	soapy	clean	clean	clean	clean
hands wet	dry	wet	wet	wet	wet	wet	wet	wet	dry
water	off	on	off	off	on	on	off	off	off
Prompt		"put some soap on your hands"		"turn the water on now"				"use the towel on your right"	

Figure 3: Example handwashing sequence. The user independently turns the water on and wets their hands (up to frame 87), at which point the system prompts them to use the soap. The user turns the water off before using the soap, so a second prompt is given to turn the water on again at frame 316. The user then manages to finish washing their hands, but stalls before drying them, so the system issues one last prompt for them to dry their hands at frame 726.
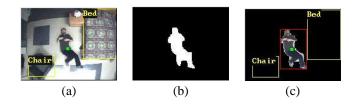

Figure 4: Example fall detection. (a) original image showing known regions (b) silhouette extracted using adaptive background subtraction (c) segmented image showing fall and regions. The size, shape and location of the person in (c) signals a fall.

denly increases, this may be an indication of the onset or recurrence of a health condition, and the transition function will assign a higher probability to this condition as a result of the observations of increased bathroom trips. Similarly, the presence of the person in an unusual location at some time of the day may also be an indication of a changing state of health. The responses to changes in these attitude variables may be to schedule a visit to a doctor, or to call a family member. The second function of the system is emergency response in the case of a fall. In this case, it is not the *attitude* variables that are important, but the *task* variables directly. The system's response will be to initiate a dialogue with the person to uncover more information, followed by possible calls to emergency response teams.

For both of these functions of the system, speech acts will be used to carry on dialogues in cases where the system attempts to uncover additional information from the user, such as whether they need help, etc. This requires some form of dialogue management encoded in the dynamics, which is an active area of research in POMDP modeling (Roy, Gordon, & Thrun 2003; Zhang *et al.* 2001; Williams, Poupart, & Young 2005).

Reward Rewards are given for calling for help if undesirable situations are encountered (e.g. unrecoverable falls, decline in health levels, etc.)

A preliminary version of this system has been imple-

mented (Mihailidis *et al.* 2005) and is currently being tested. An example of the fall detection component is shown in Figure 4. In this preliminary work, we simply engineer a policy of action without specifying the POMDP model, so there is no learning. Further, the health monitoring and emergency response systems are implemented separately. Our current work is to combine these into a single POMDP model which can be learned from data.

Wheelchair Mobility

Many older adults face various impairments and disabilities that result in their mobility being compromised. Furthermore, many of these people lack the strength to manually propel themselves, and require powered wheelchairs. However, powered wheelchairs are not appropriate for older adults with a cognitive impairment, such as dementia or Alzheimer's disease, as they do not have the cognitive capacity required to effectively and safely manoeuvre the wheelchair. In addition, their sometimes aggressive and unpredictable behavior makes wheelchair use unsafe for both themselves and others sharing the environment. As a result, many care institutions (e.g. hospitals, long-term care facilities, etc.) have implemented policies that restrict driving for reasons of safety, especially for residents with cognitive impairment.

Reduced mobility results in reduced quality of life. The combination of social isolation, limited life space and choice, learned dependence (e.g. requiring someone to push a manual wheelchair), frustration, and limited autonomy likely contributes to symptoms of depression and exacerbation of cognitive impairment and undesirable behaviors. It should also be noted that this chain reaction of symptoms resulting from reduced mobility are also observed in other user groups beyond older adults (e.g. disabled children, adults with traumatic brain injury, etc.), thus broadening the scope of these problems and requirements from potential solutions.

In this case, our POMDP model implements a mixed initiative controller for the powered wheelchair. Such ideas have been explored before (Yanco 1998; Nuttin *et al.* 2001). The idea is to let the user control their wheelchair independently until potentially dangerous situations arise, at which point the system will attempt to modify or stop the user's

controls. The system can also act in a more passive way, issuing verbal prompts or reminders if unusual or detrimental activities are noticed.

Task The wheelchair control problem can be characterised using a map of the environment in which the wheelchair operates, and maintaining the location and velocity of the wheelchair within the map. The map we use is an occupancy grid (Moravec & Elfes 1985; Murray & Little 1998), which estimates the presence of an obstacle at each 2D location $\{x, y\}$ in the reference frame of the (moving) wheelchair or in a global reference frame. Other task states may also include different aspects of the user's situation, such as their schedule, much like in the health monitoring POMDP described above.

Attitude The cognitive state of the user will have a significant impact on their ability to control a powered wheelchair, and so we will include user alertness. User frustration (with the wheelchair control) could also be included here.

Actions The actions of the system will be to modify the user actions (their behaviors), such as restricting movement in a certain direction, or modifying the user's control signal. Actions will also include audio prompts or alerts, since restricted movement on its own can cause confusion which can be mitigated by appropriate verbal cues.

Observation The behaviors of the user are their driving actions on the wheelchair (use of the control joystick, for example). They are inferred from the joystick outputs (control to the wheelchair motors). The map and location variables are estimated simultaneously from stereo vision measurements (Murray & Little 1998; Elinas 2005).

Dynamics The transition function characterises the effects of user and system controls on the map, location and velocity. The dynamics, when used to compute a policy, will give rise to implicit safety envelopes and times to collision in terms of the actions of the system and user. For example, if an obstacle suddenly appears in front of a moving wheelchair, the system can predict that the only action to avoid a collision is a hard braking one. The dynamics can also encode the larger scale patterns of a person's movement, such as their schedule, etc, as in the health monitoring case described above.

Reward The reward function will assign a large penalty for the wheelchair coming into contact with any occupied cell in the map, but a positive reward for the wheelchair responding to the user's command (e.g. from a joystick). The control policy will therefore optimise satisfaction of user's needs for mobility (their control commands) with the requirements for safety in the wheelchair's location. Other rewards can be given for keeping the user on a schedule, for example.

We have implemented a preliminary version this system that uses local occupancy grid maps and a very simple collision detection/avoidance system[1]. The wheelchair and sen-
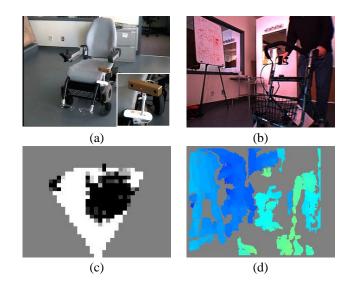

(a)   (b)

(c)   (d)

Figure 5: Wheelchair collision avoidance (a) Nimble Rocket^{TM} wheelchair and Bumblebee^{TM} stereo vision camera (inset) (b) wheelchair camera view (c) occupancy grid map - the wheelchair is at the bottom center - white, black and gray cells are unoccupied, occupied and unknown (d) stereo image - brighter is closer .

sor, and an example of a collision detection is shown in Figure 5[2]. Again, we have simply engineered a policy for this case, but our current work is focussed on implementing the system as POMDP and using it to derive policies of action. Methods for global mapping and Monte-Carlo localisation (Elinas 2005) are also being investigated.

Conclusion and Future Work

We have described a general model for cognitive assistive technologies interacting with users. The model deals with uncertainty in action outcomes and in sensor measurements in a principled way, is able to learn and adapt over time to users and tasks, and avoids the need for extensive re-engineering for new tasks by defining a framework for modeling assistive technology. We showed three examples of the model's application to technologies for wellness. Our current work is involved with the implementation and testing of versions of the models we have described above.

A direction of interest is the combination of the models we have presented above in a single, hierarchical framework. Hierarchical POMDPs have already been described in the literature (Theocharous, Murphy, & Kaelbling 2004), and we believe that similar concepts can be applied here. For example, while the health monitoring POMDP will take care of the gross movements of a user through their home, it can hand over control to the handwashing assistant POMDP when the user enters the washroom, or to the wheelchair controller when the user climbs into their mobility device.

---

[1]This work is in collaboration with the Robot Partners group at the University of British Columbia.

[2]Also see www.cs.toronto.edu/~jhoey/wheel

References

Asada, H.; Shaltis, P.; Reisner, A.; Rhee, S.; and Hutchinson, R. 2003. Wearable CV sensors. *IEEE Engineering in Medicine and Biology Magazine - Special Issue on Wearable Sensors/Systems.*

Bengio, Y., and Frasconi, P. 1996. Input-output HMMs for sequence processing. *IEEE Trans. on Neural Networks* 7(5):1231–1249.

Boger, J.; Hoey, J.; Poupart, P.; Boutilier, C.; Fernie, G.; and Mihailidis, A. 2005a. A planning system based on Markov decision processes to guide people with dementia through activities of daily living. To appear in *IEEE Trans. on Information Technology in Biomedicine.*

Boger, J.; Poupart, P.; Hoey, J.; Boutilier, C.; Fernie, G.; and Mihailidis, A. 2005b. A decision-theoretic approach to task assistance for persons with dementia. In *Proc. IEEE Int'l Joint Conf. on Artificial Intelligence*, 1293–1299.

Dearden, R.; Friedman, N.; and Russell, S. 1998. Bayesian Q-learning. In *Proc. Nat'l Conf. on Artificial Intelligence*, 761–768.

Elinas, P. 2005. Sigma-MCL: Monte-Carlo localization for mobile robots with stereo vision. In *Proc. Robotics: Science and Systems.*

Guralnik, V., and Haigh, K. Z. 2002. Learning models of human behaviour with sequential patterns. In *AAAI-02 Workshop on Automation as Caregiver*, 24–30.

Hamid, R.; Huang, Y.; and Essa, I. 2003. Argmode - activity recognition using graphical models. In *Proc. CVPR Workshop on Detection and Recognition of Events in Video.*

Hoey, J., and Little, J. J. 2004. Value-directed learning of gestures and facial displays. In *Proc Intl. Conf. on Computer Vision and Pattern Recognition.*

Hoey, J., and Poupart, P. 2005. Solving POMDPs with continuous or large discrete observation spaces. In *Proc. Int'l. Joint Conf. on Artificial Intelligence*, 1332–1338.

Hoey, J.; Poupart, P.; Boutilier, C.; and Mihailidis, A. 2005. Semi-supervised learning of a POMDP model of patient-caregiver interactions. In *Proc. IJCAI Workshop on Modeling Others from Observations*, 101–110.

Lee, T., and Mihailidis, A. 2005. An intelligent emergency response system. *Journal of Telemedicine and Telecare*. in press.

Liao, L.; Fox, D.; ; and Kautz, H. 2004. Learning and inferring transportation routines. In *Proc Nat'l Conf. on Artificial Intelligenc (AAAI '04)*, 348–353.

LoPresti, E. F.; Mihailidis, A.; and Kirsch, N. 2004. Assistive technology for cognitive rehabilitation: State of the art. *Neuropsychological Rehabilitation* 14(1/2):5–39.

Lovejoy, W. S. 1991. A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research* 28:47–66.

Mihailidis, A., and Fernie, G. R. 2002. Context-aware assistive devices for older adults with dementia. *Gerontechnology* 2:173–189.

Mihailidis, A.; Barbenel, J. C.; and Fernie, G. 2004. The efficacy of an intelligent cognitive orthosis to facilitate handwashing by persons with moderate to severe dementia. *Neuropsychological Rehabilitation* 14(1-2):135–171.

Mihailidis, A.; Tam, T.; McLean, M.; and Lee, T. 2005. An intelligent health monitoring and emergency response system. In *Proc. ICOST 2005.*

Mihailidis, A.; Carmichael, B.; and Boger, J. 2004. The use of computer vision in an intelligent environment to support aging-in-place, safety, and independence in the home. *IEEE Trans. on Information Technology in Biomedicine* 8(3):1–11.

Montemerlo, M.; Pineau, J.; Roy, N.; Thrun, S.; and Verma, V. 2002. Experiences with a mobile robotic guide for the elderly. In *Proc. of the AAAI Nat'l Conf. on Artificial Intelligence*. Edmonton, Canada: AAAI.

Moravec, H., and Elfes, A. 1985. High-resolution maps from wide-angle sonar. In *Proc. IEEE Int'l Conf. on Robotics and Automation.*

Murray, D., and Little, J. 1998. Using real-time stereo vision for mobile robot navigation. In *Proc. of the IEEE Workshop on Perception for Mobile Agents.*

Mynatt, E.; Essa, I.; and Rogers, W. 2000. Increasing the opportunities for aging in place. In *ACM Conf. on Universal Usability.*

Nguyen, N. T.; Bui, H. H.; Venkatesh, S.; and West, G. 2003. Recognising and monitoring high-level behaviours in complex spatial environments. In *Proc. CVPR.*

Nuttin, M.; Demeester, E.; Vanhooydonck, D.; and Brussel, H. V. 2001. Shared autonomy for wheel chair control: Attempts to assess the user's autonomy. In *Autonome Mobile Systeme 2001, 17. Fachgespräch*, 127–133. London, UK: Springer-Verlag.

Oliver, N.; Horvitz, E.; and Garg, A. 2002. Layered representations for human activity recognition. In *Proc. of Int'l Conf. on Multimodal Interfaces.*

Poupart, P., and Boutilier, C. 2004. VDCBPI: an approximate scalable algorithm for large scale pomdps. In *Advances in Neural Information Processing Systems 17 (NIPS)*, 1081–1088.

Roy, N.; Gordon, G.; and Thrun, S. 2003. Planning under uncertainty for reliable health care robotics. In *Proc. of the Int'l Conf. on Field and Service Robotics.*

Rudary, M.; Singh, S.; and Pollack, M. E. 2004. Adaptive cognitive orthotics: Combining reinforcement learning and constraint-based temporal reasoning. In *Proc. 21st Int'l Conf. on Machine Learning (ICML 2004).*

Rybski, P. E., and Veloso, M. M. 2004. Using sparse visual data to model human activities in meetings. In *Proc. IJCAI Workshop on Modeling Other Agents from Observations (MOO 2004).*

Theocharous, G.; Murphy, K.; and Kaelbling, L. 2004. Representing hierarchical POMDPs as DBNs for multi-scale robot localization. In *Proc. Int'l Conf. on Robotics and Automation (ICRA 2004).*

Walker, W.; Lamere, P.; Kwok, P.; Raj, B.; Singh, R.; Gouvea, E.; Wolf, P.; and Woelfel, J. 2004. Sphinx-4: A flexible open source framework for speech recognition. Technical Report SMLI TR2004-0811, Sun Microsystems, Inc.

Williams, J.; Poupart, P.; and Young, S. 2005. Factored partially observable Markov decision processes for dialogue management. In *Proc. IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, 76–82.

Wren, C. R.; Azarbayejani, A.; Darrell, T.; and Pentland, A. 1997. Pfinder: Real-time tracking of the human body. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 19(7):780–785.

Yanco, H. A. 1998. Wheelesley, a robotic wheelchair system: Indoor navigation and user interface. In Mittal, V. *et al.* eds., *Lecture Notes in Artificial Intelligence: Assistive Technology and Artificial Intelligence*, 256–268.

Zhang, B.; Cai, Q.; Mao, J.; and Guo, B. 2001. Planning and acting under uncertainty: A new model for spoken dialogue system. In *Proc. of Uncertainty in Artificial Intelligence*, 572–579.