The Use of Computer Vision in an Intelligent Environment to Support Aging-in-Place, Safety, and Independence in the Home

Alex Mihailidis, Brent Carmichael, and Jennifer Boger

***Abstract*—This paper discusses the use of computer vision in pervasive healthcare systems, specifically in the design of a sensing agent for an intelligent environment that assists older adults with dementia during an activity of daily living. An overview of the techniques applied in this particular example is provided, along with results from preliminary trials completed using the new sensing agent. A discussion of the results obtained to date is presented, including technical and social issues that remain for the advancement and acceptance of this type of technology within pervasive healthcare.**

***Index Terms*—Activities of daily living (ADL), computer vision, dementia, intelligent homes, older adults, pervasive computing.**

I. Introduction

A. Scope of Paper

THIS PAPER will discuss a sensing agent being developed as part of an intelligent environment that can support and guide older adults with dementia during the completion of activities of daily living (ADL). Through the discussion of this specific example, the goal of this paper is to start a dialogue within this research community on the important issues surrounding the use of *computer vision within pervasive healthcare*. This discussion will include technical aspects of such systems as well as the ethical and social implications of using vision, both in this example and in pervasive healthcare in general.

B. Aging-in-Place

Older adults constitute the fastest growing population group in North America, Europe, and Asia. As such, finding ways of supporting older adults who wish to continue living independently in their own homes, as opposed to moving to a long-term care facility, is a growing social problem [1], [2]. Achieving the goal of "aging-in-place" is becoming increasingly difficult as more older adults are living alone in their homes (often in rural areas) and as the prevalence of cognitive disability such as dementia increases.

Dementia is characterized by a sustained decline in cognitive function and memory [3]. An older adult with dementia is often unable to independently complete ADL, such as using the washroom or getting dressed, because he/she is unable to remember the proper sequence of tasks that must be completed [4]. Difficulty in caring for this population at home may result in the caregiver (often a family member) not being able to cope, and the affected person being taken out of his/her own home and placed in a care facility. As a result, the objective of aging-in-place is lost.

It has been hypothesized that through the careful placement of technological support, these difficulties can be reduced and aging-in-place can be preserved and enhanced [1], [2].

C. Application of Pervasive Computing

The goal of pervasive computing is to integrate information and computing into the everyday physical world, such that this technology is available to everyone in any context. It is envisioned that this concept of "computing everywhere" can be applied to support older adults in their own environments. Environments and homes that can intelligently aid in caregiving could play a very significant role in enhancing the ability of older adults to remain in their own homes. Such a system would be able to assist a user with specific ADL through automated prompting, assist with other activities such as environmental control and medication adherence, and actively monitor and ensure the health and safety of the person. New pervasive technology that can perform these functions could help to reduce the burden of care on family members and on the healthcare system.

A central theme in pervasive computing is building rich predictive models of human behavior from sensor data, enabling the environment to be aware of the activities performed within it. Using an accurate model, the environment can first determine the actions and intent of the user and then make considered decisions about the type of assistance to provide. As such, the sensing agent comprises an integral part of any system that is being developed. The design of the agent is especially important when used in a potentially sensitive application, such as a system that monitors activity within a private setting (e.g., a home) and when providing assistance to vulnerable populations such as older adults with dementia.

II. Rationale for Using Vision

In deciding the type of sensing agent to use in an intelligent environment, or any type of pervasive healthcare system, it is important to understand the user population and its requirements. A better understanding of these requirements will help in

Manuscript received November 28, 2003; revised February 13, 2004. This work was supported in part by a grant from the Alzheimer Society of Canada and Intel Corporation.

A. Mihailidis is with the Department of Occupational Therapy, University of Toronto, Toronto ON M5G 1V7, Canada (e-mail: alex.mihailidis@utoronto.ca).

B. Carmichael and J. Boger are with the Centre for Studies in Aging at Sunnybrook & Women's College Health Sciences Centre, Toronto ON M4N 3M5, Canada (e-mail: brent.carmichael@utoronto.ca; jen.boger@utoronto.ca).

Digital Object Identifier 10.1109/TITB.2004.834386

determining the types of functions that the agent needs to perform, and hence, the type of data to be collected.

Research is currently being conducted to determine the functions that an intelligent environment needs to perform in order to support aging-in-place and to be useable by a wide range of users. For example, Morris and Lundell completed a study to help uncover the needs resulting from cognitive impairment in older adults that can be addressed through home computing technologies. They identified several features that such technology should be able to assist an older adult to overcome, including assisting with remembering, management of everyday life, and social connectedness [5]. Along similar themes, Mihailidis and Fernie identified two key functions that intelligent home systems must perform: 1) automated detection of ADL completion (including prompting); and 2) ensuring the safety and health of the user [2].

Developing systems that can perform these functions requires the ability to autonomously determine the actions and context of the user. The ability for an in-home system to automatically identify the person of interest and then correctly infer information such as location, activity, and intent of the user is critical to delivering the support that is needed [5]. These sensors need to be unobtrusive, embedded seamlessly into the environment (i.e., they must not alter the existing environment), and they must be easily generalizable so that the system can be used for more than one task [5], [6]. Furthermore, they must not place any further cognitive load on the user by requiring input or effort from the older adult (e.g., remembering to wear the device or providing a response) [2].

There is a wide variety of sensing technologies that could be used in this application. For example, several researchers (e.g., [7] and [8]) have investigated the use of switches, sensors, and motion detectors (similar to those used in common home burglar alarm systems), to collect data about the completion of ADL in a home. Other projects have also used these types of hardware to infer information from the collected data and make decisions. For example, the Bath Institute of Medical Engineering has been developing a smart home for older adults with dementia, called the Gloucester Smart Home. Using similar sensing devices, a bath monitor checks on the water level in the bath and the water temperature and provides a customized reminder to the user to turn off the taps [9]. Another example is work by Mihailidis *et al.* who tracked user actions during handwashing with switches and motion sensors that monitored whether the taps were on or off, when the soap was being used, and when the person's hands were inside the sink basin. These data were used by the system to play simple automated reminders to the person if a step was missed [10].

Switches and motion detectors provide reliable data; however, they often result in incomplete information about the user and environment. Schiele and Koile stated that although positional data is a rich source of information, often more complex and complete information is needed to best describe the user's context [11]. For example, a motion detector will detect that a person is in the kitchen, but cannot tell where in the kitchen the person is located, whether the person is standing, sitting, or lying down, or who the person is. This extra information may be important in the detection of falls or other accidents. In order to determine this additional type of information more sensors would be required, which often results in a cluttered and obtrusive environment. Furthermore, since these types of sensors cannot determine the identity of the person, they are unreliable when more than one person lives in the environment or if the person has a pet. Finally, the use of such sensors may not be easily transferred to other tasks and ADL, thus, a new set of hardware and configurations would be required.

Some researchers have used sensing hardware that must be worn on the person. One of the most common systems that use these types of sensors and hardware is the telephone-based personal emergency response system (PERS), which consists of the subscriber wearing a manually activated help button as a necklace or wristband, and a home communicator that is connected to a residential phone line [12]. Remote health monitoring devices have also been developed that measure and track various physiological parameters, such as pulse, skin temperature, and blood pressure [13]. They require the user to wear the device at all times, and/or to manually take the required measurements and enter the data into the system, which then automatically transmits to an evaluation and emergency service station [14], [15]. Many of the PERS and physiological-based monitoring systems are inappropriate, obtrusive, and difficult for an older adult to operate. Not only do these systems require effort from the user, sometimes with long training times in order for the person to learn how to use the required features, they also may become ineffective during more serious emergency situations (e.g., the person has a stroke). More commonly, older adults tend to forget to wear the devices or lose them.

More recently, the use of radio-frequency identification devices (RFIDs) has increased in identifying different users and tracking their actions. For example, Wilson and Atkeson use RFID tags to identify occupants entering and leaving an environment [16]. Intel's Guide Project allows a user to garner information about RFID-tagged objects they touch while wearing an antenna-embedded glove [17], [18]. Other groups have investigated RFID in identifying people, storing medical history data, and automatic drug delivery [19], [20]. The use of RFID tags presents some of the same limitations as other on-person hardware. They require the user to continually wear the sensor, and often require several sensors to be placed throughout the user's environment. Also, a new set of sensors and placements of the devices are required for different tracking tasks and ADLs. However, as described, RFID tags can be used to identify different users, a potentially important advantage of this type of technology.

It is believed that computer vision can be used to develop a sensing agent that overcomes the limitations of these types of sensing hardware. Vision can be used not only to obtain positional data, but also to collect more in-depth information in order for the system to build more accurate models of the user and the environment. For example, such an agent could be used to determine more accurately the person's location within the environment, any unusual actions, and distinguish the user of interest from other people or pets sharing the same space. A vision-based system can be used to not only track gross movements of the person, but to also track the fine motor movements (including gestures) required to monitor the completion of ADL and other tasks. Finally, markers or devices do not need to be worn on the person to collect these data, which makes using the technology less burdensome for the user. However, issues surrounding the use of vision need to be addressed in order for such systems to become more widely accepted and used. These include technical issues such as the development and optimization of image processing algorithms, as well as ethical and social issues such as those concerned with the privacy of users and the use of vision in the home (which are discussed later in this paper).

III. Relevant Work Using Vision

In recognition of the potential benefits of using a vision system, several new research projects in the area of intelligent environments, specifically within the context of healthcare for older adults, are currently developing vision-based sensing agents. For example, researchers at the Georgia Institute of Technology are involved in the Aware Home project, a prototype that is currently being designed as a "living laboratory" for the development of context-aware and artificially intelligent computing in support of aging-in-place [21]. The house has a wide range of sensing equipment, including video cameras and computer-vision agents to monitor the actions and activities of older adults. The goals of this hardware are to automatically and unobtrusively measure activities of the residents and provide support for their daily needs and activities [1], [21], [22]. The Agent-Based Intelligent Reactive Environment Group at MIT uses a computer-vision system to monitor the location of a person in a room in order to determine which activities are being completed, and with which objects a user is interacting [23]. Finally, McKenna *et al.* of the University of Dundee's Applied Computing group have demonstrated positive results tracking older adults using computer vision to detect falls as part of their automated home monitoring research [24].

The authors developed an early prototype of the example to be described in this paper. The system used simple computer vision and artificial intelligence (AI) to learn to associate hand positions (two-dimensional coordinates) with specific handwashing steps (e.g., turning on the water, using the soap), and to adjust the system's parameters and cueing strategies to meet the changing needs and preferences of each user [6]. Clinical trials involving ten subjects with moderate-to-severe dementia showed that the number of handwashing steps completed without assistance from the caregiver increased by approximately 25% when the device was used [25]. Although the computer-vision agent used in this prototype showed some success, these results have identified several remaining challenges. In particular, the agent required a bracelet to be worn by the user for tracking, which is obtrusive, bothersome, and may be removed by the user.

These research projects have begun to investigate whether or not computer vision can play a role in the areas of intelligent environments and pervasive healthcare. However, to date, these systems have limitations, which may make them inappropriate for use in an intelligent environment that is capable of providing care to an older adult. These systems were developed primarily to monitor gross body movements within an environment and to provide general assistance, such as automatic control of lights or general reminders and prompting. Systems that do focus on ADL monitoring require very specific user and environmental conditions to work appropriately. For example, the person must be wearing long sleeves and strict environmental controls such as clear repeatable backgrounds must be used (e.g., background objects must always be in the same locations). Obviously, any system that will be used in the home cannot assume such ideal conditions.

In order to continue the advancement of vision-based sensing agents, this work needs to be built upon through the development and application of vision algorithms that can support the additional functionality that is required for an intelligent environment to support aging-in-place. Further investigations of social and ethical issues also need to be explored to be able to make any conclusions on the appropriateness of such systems.

IV. Overview of New Research

A. Research Goal

In attempt to address some of these limitations and to continue to investigate the potential of using computer vision in intelligent environments, the authors and their collaborators have been conducting research in the development of a new system for assisting older adults with dementia.

The overall goal of this research is to apply the principles of pervasive computing to develop a new and enabling technology to improve an older adult's independence, safety, and quality of life at home. Specifically, an intelligent environment is being developed that will unobtrusively monitor the actions of an older adult with dementia during a common ADL and provide in-depth and personalized assistance using verbal and visual prompting.

B. Overview of System

The new system will consist of three agents: 1) sensing; 2) planning; and 3) prompting. The sensing agent will monitor the actions of the user by determining the spatial coordinates of the person's body and hands within the environment. Once these coordinates and the step the person is completing have been determined, the system's planning agent will determine which plan (i.e., sequence of steps) the user is attempting, and whether the step being completed is appropriate. If the system detects that the user has made an error, such as completing a step out of sequence or missing a step altogether, a customized prompt to guide the user will be selected and played automatically.

At this stage of development, the system is being designed for the specific ADL example of handwashing, which includes the following steps: 1) turning on the water; 2) wetting of hands; 3) applying soap; 4) rubbing soap on hands; 5) rinsing soap off of hands; 6) turning off the water; and 7) drying hands. Note that there is more than one acceptable sequence to perform these steps, which will be taken into account by the system's planning agent (beyond the scope of this paper). Handwashing was chosen as the development platform for three reasons: 1) it is easily definable and relatively simple to model for the development of the required agents; 2) it is complex enough that it poses difficulties for older adults with dementia, which often results in the ADL not being completed properly; and 3) it is relatively safe for clinical trials, without as many concerns about privacy as other ADL. The authors have also used this ADL during the development of previous prototypes, for example, [6] and [10].

Each of these agents is currently under development, however, this paper will focus on the sensing agent and the work completed thus far. The other system agents will be the focus of future publications, as will the results of clinical trials completed with the new prototype.

C. Overview of Sensing Agent

The following example will illustrate that relatively simple and commonly used computer-vision techniques can be potentially used to develop a system that monitors home environments. It will also identify some of the key issues that need to be examined in order for vision to gain acceptance within this field of research.

A vision-based agent is being developed to track the actions of an older adult during handwashing. Using a frame-by-frame process (i.e., there is not temporal dependence), the agent uses color as the marker in order to perform two functions: 1) tracking of hand location; and 2) tracking of step-specific object locations, specifically the soap and towel. A combination of statistics-based color segmenting models and background subtraction (BGS) is used to identify these objects within the field of view. Using blob analysis and custom logic rules, the agent determines the two-dimensional $[x, y]$ coordinates of these various objects, which it uses to determine which handwashing step the user is completing.

The sensing agent was developed using a Sony SSC-DC393, 1/3-CCD color video camera (mounted directly above the bathroom vanity), and a Matrox Meteor II frame grabber installed on a 2.4-GHz personal computer. The agent and graphical user interface were developed using Microsoft Visual Basic v.6.0 and the Matrox Imaging Library v.7.0.

V. Development of Sensing Agent

A. Design Criteria

Several design criteria were outlined based on findings from previous research in order to develop a sensing agent that would be appropriate for the older adult user population and for this application, including the following.

1) The agent must be able to perform the required operations in real-time.
2) The agent must be able to track fine motor movements, such as hand movements, as well as gross overall body movements.
3) The agent must be able to work in real, nonideal conditions, such as in a washroom, and not be dependent on the appearance of the user (e.g., types of clothing, etc.).
4) The agent must be able to operate against a dynamically changing and potentially cluttered background.
5) The user must not be required to wear any type of marker or device.
6) The agent must be easily generalizable for use in different tasks (beyond the initially chosen ADL of handwashing).

B. Color-Based Tracking

A brief introduction to color-based tracking is necessary; however, a detailed look at the theoretical aspects of this type of sensing agent is beyond the scope of this paper. The reader is referred to the references in this section for more details.

In order to track human motions without applying markers, the agent must rely on some distinguishing natural feature such as shape, edges, or color. Color tracking has the advantages of being orientation- and size-invariant while being fast to process [26]. Therefore, the use of skin color as a nonobtrusive marker for real-time tracking systems seems to be a potentially ideal solution for this application.

Both statistics-based [27], [28] and physics-based [29] methods of segmenting skin color in digital images have been investigated for face and hand tracking. Physics-based methods are more robust since they incorporate the spectral sensitivity of the camera, the correlated color temperature values of the light source, and the spectral reflectance values for human skin. Knowledge of these parameters makes the segmentation module more reliable if the illumination changes. However, obtaining knowledge of each of these parameters is difficult, and as a result, the majority of color-based tracking systems to date have only implemented statistical models.

There are several different types of statistics-based methods for segmenting skin color. Studies have shown that skin tones differ primarily in their intensity (the strength or vividness of a color) while their chromaticity or hue (the quality of a color as determined by its dominant wavelength) remains relatively unchanged [27]. As a result, after testing many color models Terrillon *et al.* and Martinkauppi determined the normalized color coordinate, or normalized chromatic coordinate (NCC) model to be the most effective for skin segmentation since this model produced the smallest area for skin chromaticities [27], [30]. The NCC model involves dividing each of the individual red-green-blue (RGB) components of a given color by its intensity, which is equivalent to the sum of these components, for example, $r = R/(R+G+B)$. This normalizes each of the component values with respect to the overall intensity of the color allowing for comparisons of different colors to be made purely on a chromatic basis. Another statistics-based method of interest in the development of a color-based sensing agent is the hue, saturation, intensity (HSI) color model. The HSI model decouples the intensity component from the color-carrying information (hue and saturation) in a color image. As a result, this model is an ideal tool for developing image processing algorithms based on color descriptions that are natural and intuitive to humans [31]. This is particularly useful when the objects to be tracked need to be easily defined by the system. For example, the system can easily be set to track all objects in the field of view of a specific color.

C. Tracking of Hand Locations

The sensing agent determines the location of new objects introduced into the field of view via the BGS method. BGS is one of the simplest and quickest methods of optical filtering.

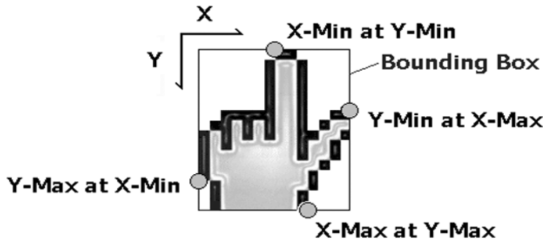

Fig. 1. Definition of contact points. Contact points indicate the four to eight possible intersection points of the blob's perimeter with its bounding box. In this case, four contact points for each blob are used.

Grayscale values of each pixel are compared to an initial reference image that was recorded before the object of interest entered the field of view. If the grayscale value of a particular pixel differs from the reference value by more than 10% (determined via trial and error under various environmental conditions), a new object must exist at this location.

The NCC method is then used to locate all skin-colored pixels of these new objects creating a corresponding binary image. As described, pixel values of each image are converted to r-g components. If the summation of the r and g values determined from the $[R, G, B]$ pixel value fall within the range of 0.695 and 0.750 (found empirically and taken from literature [29]), then this pixel is recognized as skin and the corresponding binary image pixel value is set to "1."

A blob analysis algorithm identifies each of the skin-colored objects and ignores all but the four largest blobs that are greater than 200 pixels in size (within a 160×120 pixel frame). Up to four blobs are of interest, taking into account the user may be wearing a watch or bracelet on either wrist that could confuse the agent by dividing the arm into two distinct objects. Location and orientation information from each blob is then used to determine which pair of possible blobs corresponds to the user's hands regardless of size, thus, users with bare arms or long sleeves can both be tracked effectively. The palm centers of these two blobs are tracked and their two-dimensional $[x, y]$ coordinates are available for use by the planning and prompting agents.

The selection of which blob contains the user's hand is based on the location of each blob's contact points, a feature contained within the MIL development software, as illustrated in Fig. 1.

Based on the location of these points, the direction the arm is pointing as well as which disconnected blobs are related can be determined. This is accomplished according to the following rules that were developed.

1) It is assumed the user will not lead with their elbow and that blobs furthest away from the user contain the hands.
2) If Blob A's bottommost contact point (X-Max at Y-Max) is very near Blob B's topmost contact point (X-Min at Y-Min), then Blobs A and B are related and Blob A contains the user's hand (Fig. 2).
3) If Blob C's rightmost contact point (Y-Min at X-Max) is very near Blob D's leftmost contact point (Y-Max at X-Min), then Blobs C and D are related and if Blob C is furthest from the user then it contains user's hand (Fig. 3), otherwise, Blob D contains the user's hand (Fig. 4).

Fig. 2. Determining the hand's location given two vertically stacked blobs. The first frame represents the actual image from the video sequence, the second is the binarized image after the NCC model has been applied, and the third is a representation of the palm centers after blob analysis has been performed.

Fig. 3. Determining the right hand's location given two horizontally stacked blobs. The three frames illustrate the same stages of processing described in Fig. 2.

Fig. 4. Determining the left hand's location given two horizontally stacked blobs. The three frames illustrate the same stages of processing described in Fig. 2.

4) The only blobs that are tracked are those containing hands.

If the user is not wearing a wristwatch or bracelet on one or both of their arms, then the contact points are used to determine which end of the long trapezoidal-shaped blob contains the user's hand. Again it is assumed that the user will not lead with their elbow and that blobs or regions furthest away from the user's body contain their hands. In this case, it is also further assumed that if the major (long) axis of the arm is parallel to the counter's edge, then the hand is located near the blob's rightmost end (for the case of the left arm), and that the hand is located near the blob's leftmost end (for the case of the right arm). Right and left are determined based on the location of the other blob, or group of blobs. If only one blob exists, then right and left are determined based on which half of the frame the blob's center of gravity resides.

Another special feature of the sensing agent is the speed filter, which removed any intermediate points that were tracked when the user moved their hand from one task region to the next, such as from the soap dish to sink basin. In doing so, irrelevant data was not passed to subsequent portions of the device. This decreased the possibility of errors and reduced the processing time of the whole device. Any movement greater than 100 pixels per second for a standard 640×480 pixel image was ignored. This factor was scaled for smaller image sizes and was chosen based on a similar filter that was used by Mihailidis *et al.* [6] in their second prototype.

D. Tracking of Step-Specific Objects

The sensing agent determines the initial locations of the objects that are used during handwashing, such as the soap and towel, and continually tracks their locations and the user's interactions with them. This avoids the need for the step objects

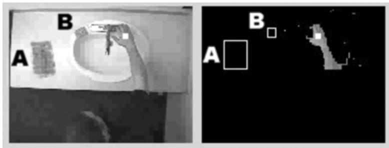

Fig. 5. Tracking both step-specific object and hand positions. The first frame shows the actual image from the frame, while the second frame illustrates the objects and hand being tracked after the HSI model (for the objects), NCC model (for the hand), and blob analysis (both types of objects) have been applied.

and the background to look identical each time the person uses the overall system.

The agent is first "taught" the colors of the soap and towel by highlighting (using manually drawn bounding boxes) these step specific objects in an initial photo of the vanity surface at system setup. During tracking of the new objects that are not skin-colored, those that resemble the soap and towel in color (based on HSI) are also tracked by blob analysis, determining the spatial coordinates of these objects. Fig. 5 illustrates the agent tracking object and hand positions. The towel is marked with the label "A," and the soap with the label "B."

E. Determining the ADL Step Being Completed

Once the spatial coordinates of the hands have been determined, the agent determines which step is being completed through a training process that allows the system to relate hand positions to steps. This is accomplished using algorithms similar to those used in previous prototypes by the authors. Readers are referred to publications such as [6] for more details on these algorithms.

In addition to using spatial coordinates to determine which step is being completed, the tracking of step-specific objects is also used. According to each object's spatial history and the location of the user's hands relative to those objects, the agent determines whether or not an object has been used. Three rules are currently being developed (and tested) to determine this information.

1) If the object becomes occluded or changes shape while the user's hand is nearby, then the object is likely being used;
2) If the object's position changes along with a change in position of the person's hand, then the object is being used;
3) If the user's hand moves away from the object and the object's position and shape remain unchanged, then the object is not being used.

VI. Preliminary Testing of Sensing Agent

A. Overview of Experiments and Data

Preliminary data was collected through the completion of several experiments conducted under controlled settings. These trials were conducted in order to determine the efficacy of the sensing agent developed thus far and to identify any remaining challenges for future development. The setup for these trials included a mock-up of a washroom, in which several subjects simulated the required handwashing steps. A white glossy counter top and sink with silver faucets was used along with a light-blue towel and green bar of soap.

The primary experiment was to verify the accuracy and repeatability of the agent (i.e., how well it was able to track and define the hand positions associated with each ADL step). Six subjects of varying skin tone were asked to move their hands to five locations marked on the vanity countertop, in any particular order (each subject was instructed to "wash their hands" as they normally would) as the agent tracked their hand locations. These locations were selected to simulate handwashing (i.e., the taps, under the faucet, the soap dish, and the towel). This procedure was repeated three times by each subject, producing 3970 frames of data. These trials were conducted for tracking only one hand and tracking both hands. The results of these trials are illustrated in Figs. 6 and 7. The locations of the actual objects are annotated in these figures in order to provide a comparison between the actual and calculated locations.

On average, the agent correctly identified 83% of a user's skin-colored area, which was determined by manually segmenting skin-colored regions in the pre-processed image and comparing the number of actual skin-colored pixels to the number calculated by the agent. The agent correctly identified the users' hands and palm centers 99.5% of the time, with one miss and no false positives (based on statistics calculated for a random sample of approximately 200 frames of data from the overall data set). The processing speed of this method was calculated to be approximately ten frames per second (Fr/s). This sampling rate was deemed acceptable for tracking human motions since volitional movements normally occur at approximately 3 Fr/s, and a tracking system should be able to operate at two to three times this value [32].

Finally, the efficacy of the agent under different lighting conditions was tested. Previous studies using the NCC model note that lighting changes can affect the r-g space skin tone values. A slight change in the agent's ability to correctly recognize skin-colored objects was observed between the use of incandescent and florescent lighting. Also, more pixel noise was observed under incandescent lighting. However, it was concluded these changes were not significant enough to alter the effectiveness of the agent and overall system.

B. Discussion of Results

The results of these experiments and general observations of agent performance were very positive. As illustrated in Figs. 6 and 7, definitive clusters of data for each required position were obtained, indicating high repeatability (in comparison to previous prototypes developed by the authors). With respect to the accuracy of the agent, the data clusters were approximately in the same locations of the actual positions/objects. The offset observed between the actual and calculated positions was due to the users "interacting" with the required objects with their fingertips (i.e., using grasping motions), while only their palm centers were calculated by the agent. Again, this finding was comparable to previous prototypes. This result was deemed to be acceptable for this particular application since the overall system will be trained according to these cluster locations. Thus, as long as the accuracy is good (as is the case in this example), repeatability is the most important criterion in this particular application.

Even though some simple statistics have been collected, the reliability that is required by such a system and whether the cur-

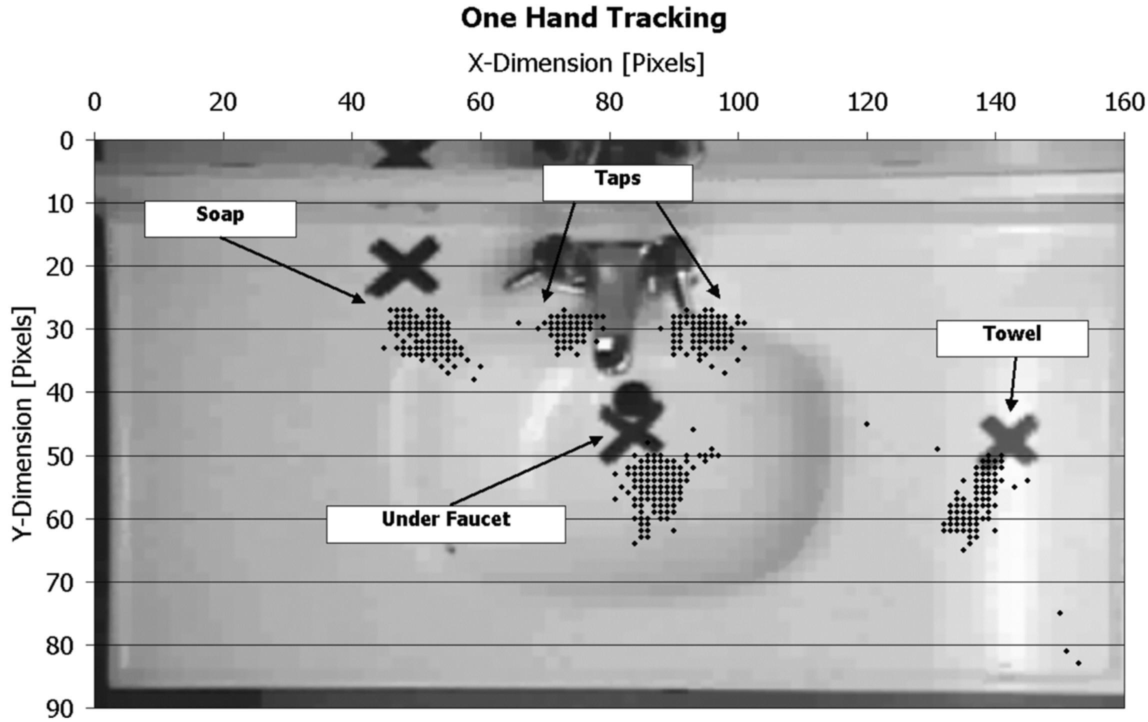

Fig. 6. Results for one hand tracking. Each cluster represents data points collected for the $[x, y]$ positions of one hand of the user as each required handwashing step is completed. The actual locations of the positions/objects represented by these clusters are annotated.

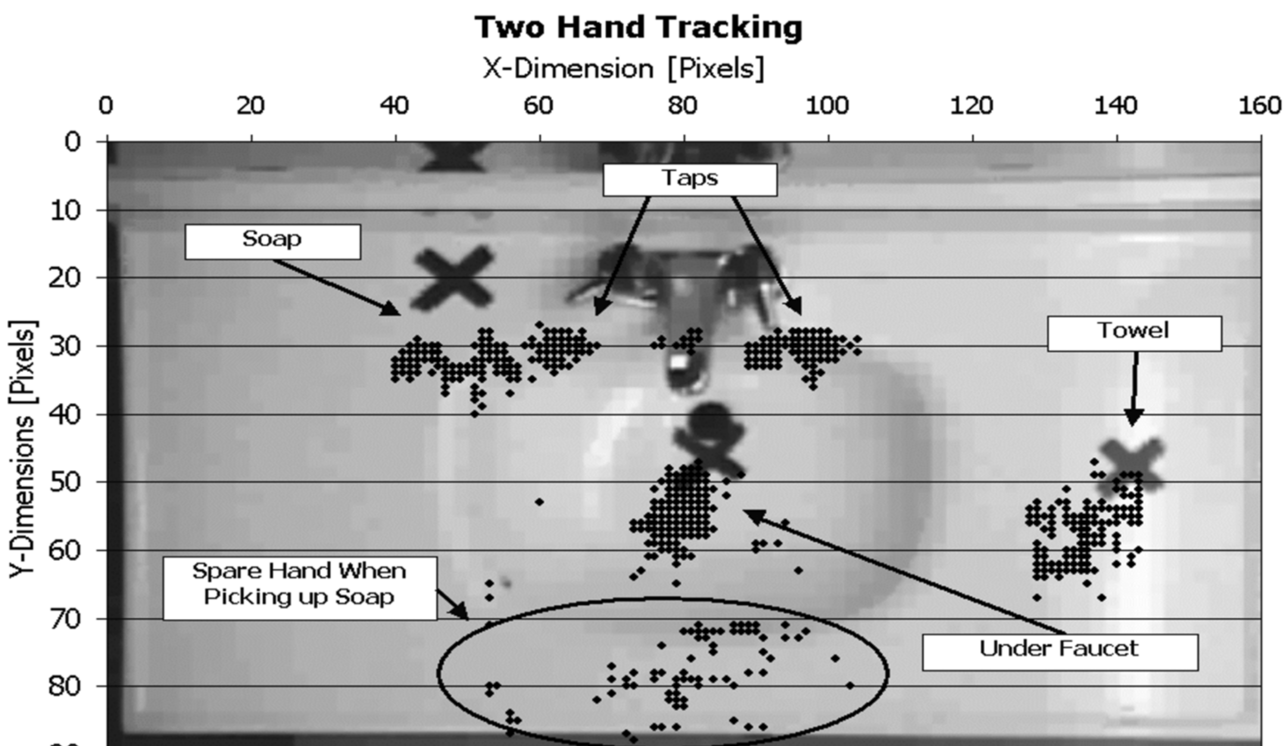

Fig. 7. Results for two-hand tracking. Each cluster represents data points collected for the $[x, y]$ positions of both hands of the user as each required handwashing step is completed. The actual locations of the positions/objects represented by these clusters are annotated. These data do not differentiate between the user's left and right hands.

rent agent meets this requirement will not be known until the overall system has been tested. Reliability of such a system (or any type of assistive technology) should not be based solely on the individual components, but rather on how the overall system performs its required functions. It is expected that the level of accuracy and repeatability observed for the sensing agent will be sufficient for subsequent portions of the overall system to determine which handwashing step is being performed, allowing for proper decisions to be made.

The majority of the design criteria for this agent were met. The agent was able to perform required operations above the recommended sampling rate for tracking human volitional movements without any type of marker worn on the user. It worked properly independent of the user's sleeve length, skin color, or arm apparel (e.g., wristwatch or bracelets), and within a semi-cluttered environment. The latter, however, needs further study in a more realistic setting (i.e., using real running water, different lighting conditions, etc.). Further testing is also required to determine whether or not the developed agent can work for ADL besides handwashing, as well as other pervasive healthcare applications. It is expected that for ADL tasks that are definable by hand movements and positions (e.g., meal preparation and cooking), this agent will work well. For more complex ADL, such as toileting or dressing, further development will likely be required.

These experiments also revealed several challenges that still remain in the development of a vision-based agent for this particular application. As illustrated in Fig. 7, with two-hand tracking, the clustering was not as well localized as when only one hand was tracked. This may partly be because the subject only focused on moving one hand at a time while the second hand drifted through other parts of the tracking space. For instance, when reaching for the soap with the left hand, the right hand often drifted toward the near edge of the sink. This has the positive effect of showing that the hands are being tracked properly, though indicates that perhaps additional logic and intelligence needs to be incorporated in this type of agent in order to distinguish between hand positions that are indicative of a step being completed, and those that are irrelevant (e.g., if one hand drifts toward an incorrect step, while the other correctly performs a task, the actions of this stray hand will have to be ignored). Also, as can be seen in Fig. 7, the location of the soap dish is less distinguishable in the two-handed cases than in the one-handed cases (Fig. 6). This may be because subjects were not concentrating on the target as well as in the one-handed trials. These data do suggest that special care will have to be taken in selecting the boundaries of "task" regions when designing the logic pathways for each step. Furthermore, as illustrated in Fig. 6, some points were plotted between the cold and hot water taps. When the hands were positioned close together but not overlapping, the sensing agent viewed these objects as overlapping due to the resolution of the vision system and the binary image noise (random pixels in the foreground of the binary image that do not correspond to skin-colored objects). Therefore, work will have to be done to improve the resolution of the computer-vision system and decrease the pixel noise without compromising the sampling rate of the agent. Finally, the agent described relies exclusively on a single two-dimensional view and, therefore, the heights (z plane) of the user's hands are unknown. This was particularly problematic near the taps, since the sensing agent could not conclusively determine if the user had actually touched them or simply held their hand above them. This raises the alternative methods of using a three-dimensional vision system, or multiple camera angles. However, previous attempts at such systems have resulted in poor performance with respect to sampling rates. This also raises the question on whether a hybrid sensing agent that uses simple switches and sensors in addition to vision may be necessary. A solution that is currently being developed by the authors and their collaborators is the use of sophisticated AI planning algorithms, in particular a partially observable Markov decision process, that would be able to take into account this uncertainty to estimate whether a particular step has been completed or not [33].

Overall, the system has proven to be successful in tracking objects of interest in a near real-time, discrete, and unobtrusive manner, demonstrating that vision may be an effective sensing technique for this particular application. In comparison to the previous agent developed by Mihailidis *et al.* that used switches and sensors to track the completion of handwashing [10], this new agent has several advantages. Specifically, the hardware required by this new agent is simpler resulting in fewer modifications to the environment, the hardware is easily adaptable from one location to another and, with more development, from one ADL to another (e.g., from handwashing to toileting). This new agent also has more opportunities to be expanded to include other types of data beyond position. For example, work is currently being planned to develop algorithms to determine hand postures and gestures in order to better define the required steps.

These preliminary results will be used to plan future work related to this specific agent and the development of the overall system. Identified improvements will be implemented into the sensing agent and further testing will be conducted (as previously outlined in this section). More importantly, future work will focus on implementing the sensing agent into the overall prompting system, which will include integrating this agent with the planning and prompting agents. Both bench testing using simulated conditions and clinical trials with actual users with dementia will be completed to determine the overall system's efficacy and reliability.

VII. Conclusion

Through the presentation of a relatively simple example of a vision-based sensing agent, this paper has raised the possibility of using computer vision to develop intelligent environments to support aging-in-place. Although the agent developed thus far only focused on a small subset of the functions that an intelligent environment/home would need to perform to support this population, the combination of these preliminary results and those obtained by other researchers provide evidence that this type of sensing agent is a viable option.

However, as identified in this work and alluded to earlier in this paper, limitations still exist in using vision in this particular application. If left unresolved, these limitations may continue to result in the relatively low rate of acceptance of such systems. The question remains then, what goals and milestones need to be achieved for computer vision to gain further acceptance in this field? Preliminary findings and observations from the example presented have helped to identify some required key areas of research.

With respect to the technical aspects of such systems, researchers need to continue to develop and apply new algorithms that allow systems to perform the functions that are required. For example, observations during preliminary testing illustrated the need for more in-depth information about the user's actions beyond positional data. The agent requires further work to ensure the generaliziability of the algorithms developed and the expansion of the agent to include not only hand tracking, but also full body tracking capabilities. Other advancements such as the use of gesture recognition and prediction of movements will also likely be required to perform some of the important functions previously identified by Morris and Lundell [5] and Mihailidis *et al.* [2]. These additional capabilities will also make this type of agent more applicable in other pervasive healthcare systems that require more in-depth data about a user. Furthermore, new algorithms need to focus on ensuring that real-time performance is maintained, which may prove to be a strong limitation of vision systems in this application. The authors' experiences in developing intelligent environments and systems for older adults have shown that the time of response is critical. If there is a significant delay between the detection of a "situation" and providing assistance, such as playing a reminder to complete a particular step, the prompt will most likely confuse the user. This is not only true when assisting older adults, but for most other types of users, especially in circumstances where the person may be working in a complex environment (e.g., in a hospital). Finally, it is becoming more important to ensure that these systems are affordable by attempting to use off-the-shelf hardware and commonly used algorithms. For example, research is being completed on the development of low-cost vision systems that use significantly cheaper universal serial bus (USB) cameras instead of the typical camera frame-grabber combination [34]. The work described in this paper has shown that relatively simple and commonly used computer-vision algorithms can be used in this application in order to develop a system that can perform the required functions while being computationally inexpensive. Developing new sensing agents that are more affordable and easier to apply will make vision more attractive for use not only in intelligent environments, but as well as in other pervasive healthcare applications.

The ethical and social implications of using vision systems requires more in-depth study. There has been a long debate about the use of monitoring systems in both private and public spaces, especially surrounding the issue of video cameras. This issue becomes even more confounded when proposing to develop such systems for use during potentially private ADL and with populations who may not be able to provide consent, such as older adults with dementia.

Preliminary research and discussions on this topic have tended to focus on the ethical issues surrounding privacy and pervasive computing, including examining issues related to the use of such technology in daily life, when should these devices be used, and who should control the use of this technology [35]. The bulk of the literature on the ethics of using video has focused on the cases of surveillance by governments and by employers [36]. With respect to the use of sensors, in particular vision systems in the home and healthcare environments, very little research has been conducted and published. As such, there is little data to support the argument either for or against the use of vision in this application. Many of these arguments have been based on personal accounts and experience and on pilot studies normally completed with small sample sizes. For the specific example of assisting older adults, McKenna *et al.* [37] found that potential older adult users were content to have a monitoring device based on visual tracking provided that only a computer analyzes the output. In addition, the authors' experiences from developing the agent described in this paper and previous prototypes that used vision have indicated that through proper education and explanation of how such agents operate (in this case, directed toward the user's caregiver, family members, etc., as the primary users have moderate-to-severe dementia), ethical concerns are reduced. These findings suggest that this type of sensing agent would be acceptable if designed so that actual video is not being viewed by an operator (i.e., the images are only used by the system to process and collect the data required, and images are not stored). In this respect, the vision system is thought of and acts as a more sophisticated type of switch or motion sensor. This notion has been the philosophy of the research and sensing agent described in this paper.

It has become clear that the issue of acceptance of vision systems (both among users and developers) is becoming more important and one that deserves more discussion. More in-depth studies need to be conducted which explore critical issues such as: determining the types of home monitoring technology that community-dwelling older adults would accept into their homes; during which ADL and tasks would such technology be acceptable; and who will view data collected from such systems and during which type of circumstance. Furthermore, the ethics involved with developing and using such systems with potentially vulnerable populations needs to be studied. For example, who makes the decisions for this population on when and how to use such technology? Questions related to the future of this work also need to be addressed, such as how will these ethical issues change in the future when such technologies are used by the current baby boomer generation? It is expected that the acceptance of this type of technology within this user population will increase, and that the ethical issues that currently surround this topic will not be as prevalent. Finally, the ethical implications of pervasive technology and healthcare in general need to be explored, such as issues related to health data transmission.

In summary, it has been proposed that computer vision can have a significant role in the development of intelligent environments to support older adults. The example presented in this paper has provided evidence that a vision-based agent has the potential to perform the functions required from such a system and that it can be used effectively to support aging-in-place. As described, vision has the potential to address several limitations of current systems developed for this application, providing several advantages over other types of sensing hardware. Furthermore, it is believed that the results and lessons learned from this specific example are relevant outside of this application. Vision can potentially have a significant role in other pervasive healthcare systems, such as medical diagnostic and health monitoring systems. However, this work has identified several limitations, both technical and ethical, that need to be resolved for vision systems to gain further acceptance in this research field. Addressing these limitations and concerns must remain an important goal in any future development efforts in order to ensure that such systems are appropriate and effective for their intended user populations.

REFERENCES

[1] E. D. Mynatt, I. Essa, and W. Rogers, "Increasing the opportunities for aging in place," in *ACM Conf. Universal Usability (CUU)*, Arlington, VA, 2000, pp. 65–71.

[2] A. Mihailidis and G. Fernie, “Context-aware assistive devices for older adults with dementia,” *Gerontechnology*, vol. 2, pp. 173–189, 2002.
[3] R. Lubinski, *Dementia and Communication*. Philadelphia, PA: B.C. Decker, Inc., 1991.
[4] J. Cockburn and C. Collin, “Measuring everyday memory in elderly people: A preliminary study,” *Age and Ageing*, vol. 17, pp. 265–269, 1988.
[5] M. Morris and J. Lundell, *Ubiquitous Computing for Cognitive Decline: Findings From Intel’s Proactive Health Research*. Seattle, WA: Intel Corp., 2003.
[6] A. Mihailidis, G. R. Fernie, and J. C. Barbenel, “The use of artificial intelligence in the design of an intelligent cognitive orthosis for people with dementia,” *Assistive Technol.*, vol. 13, pp. 23–39, 2001.
[7] M. Nambu, K. Nakajima, A. Kawarada, and T. Tamura, “A system to monitor elderly people remotely, using the power line network,” presented at the World Congress on Medical Physics and Biomedical Engineering, Chicago, IL, 2000.
[8] M. Ogawa, S. Ochiai, K. Shoji, M. Nishihara, and T. Togawa, “An attempt of monitoring daily activities at home,” presented at the World Congress on Medical Physics and Biomedical Engineering, Chicago, IL, 2000.
[9] T. Adlam, C. Gibbs, and R. Orpwood, “The gloucester smart house bath monitor for people with dementia,” in *Physica Medica, Eur. Congress of Medical Physics and Clinical Engineering*, vol. 17, 2001, p. 189.
[10] A. Mihailidis, G. R. Fernie, and W. L. Cleghorn, “The development of a computerized cueing device to help people with dementia to be more independent,” *Technol. & Disability*, vol. 13, pp. 23–40, 2000.
[11] B. Schiele and S. Antifakos, “Beyond position awareness,” presented at the 3rd Int. Conf. Ubiquitous Computing, Atlanta, GA, 2001.
[12] *The Study of Emergency Response Systems for the Elderly*, MacLaren-Plansearch, Canada Mortgage and Housing Corporation, Ottawa, 1988.
[13] F. G. Miskelly, “Assistive technology in elderly care,” *Age and Ageing*, vol. 3, no. 6, pp. 455–458, 2001.
[14] H. Asada, P. Shaltis, A. Reisner, S. Rhee, and R. Hutchinson, “Mobile monitoring with wearable photoplethysmographic biosensors,” IEEE Eng. Med. Biol. Mag., vol. 22, pp. 28–40, Mar. 2003, to be published.
[15] K. Doughty, R. Lewis, and A. McIntosh, “The design of a practical and reliable fall detector for community and institutional applications,” presented at the Telemed ’99 Conf., London, U.K., 1999.
[16] D. Wilson and C. Atkeson, *The Narrator: A Daily Activity Summarizer Using Simple Sensors in an Instrumented Environment*. Pittsburgh, PA: Carnegie Mellon Univ., 2003.
[17] D. Patterson, D. Fox, H. Kautz, and M. Philipose, “Expressive, tractable and scalable techniques for modeling activities of daily living,” presented at the 2nd Int. Workshop Ubiquitous Computing for Pervasive Healthcare Applications (UbiHealth 2003), Seattle, WA, 2003.
[18] I. R. Seattle, “Intel Research Seattle Guide Project Website,” Intel, 2003.
[19] M. Sloman, “Ubiquitous Computing for Healthcare in the Community,” Ubicare, 2003.
[20] C. J. Murray, “Implantable Chips Make First Run at Personal ID,” Embedded.com, 2003.
[21] C. D. Kidd, R. Orr, G.D. Abowd, C.G Atkeson, I.A Essa, B. MacIntyre, E. Mynatt, T.E. Starner, and W. Newstetter, “The aware home: A living laboratory for ubiquitous computing research,” in *2nd Int. Workshop on Cooperative Buildings*, 1999, pp. 191–198.
[22] D. J. Moore, I. A. Essa, and M. H. H. 3rd, “Exploiting human actions and object context for recognition tasks,” presented at the 7th IEEE Int. Conf. Computer Vision 1999 (ICCV’99), Corfu, Greece, 1999.
[23] K. Koile, K. Tollmar, D. Demirdjian, H. Shrobe, and T. Darrell, “Activity zones for context-aware computing,” presented at the 5th Int. Conf. Ubiquitous Computing (UbiComp 2003), Seattle, WA, 2003.
[24] S. J. McKenna, F. Marquis-Faulkes, P. Gregor, and A. F. Newell, “Scenario-based drama as a tool for investigating user requirements with application to home monitoring for elderly people,” presented at the HCI Int., Crete, Greece, June 22–27, 2003.
[25] A. Mihailidis, J. C. Barbenel, and G. Fernie, “The efficacy of an intelligent cognitive orthosis to facilitate handwashing by persons with moderate-to-severe dementia,” *Neuropsychol. Rehab.*, vol. 14, pp. 135–171, 2004.
[26] M. Stoerring, H. J. Andersen, and E. Granum, “Skin color detection under changing lighting conditions,” presented at the 7th Symp. Intelligent Robotics Systems, Coimbra, Portugal, 1999.
[27] “Face color under varying illumination – Analysis and applications,” Dept. Elect. and Inform. Eng., Univ. of Oulu, Oulu, Finland, 2002.
[28] S. J. McKenna, Y. Raja, and S. Gong, “Tracking color objects using adaptive mixture models,” *Image Vis. Comput.*, vol. 17, pp. 225–231, 1999.
[29] M. Stoerring, H. J. Andersen, and E. Granum, “Physics-based modeling of human skin color under mixed illuminants,” *Robot. Autom. Syst.*, vol. 35, pp. 131–142, 2001.
[30] J. C. Terrillon, Y. Niwa, and K. Yamamoto, “On the selection of an efficient chrominance space for skin color-based image segmentation with an application to face detection,” presented at the Int. Conf. Quality Control by Artificial Vision, Bourgougne, France, 2001.
[31] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
[32] D. A. Winter, *Biomechanics and Motor Control of Human Movement*, 2nd ed. Toronto, Canada: Wiley, 1990.
[33] J. Boger, A. Mihailidis, P. Poupart, and G. Fernie, “Using a POMDP controller to guide persons with dementia through activities of daily living,” presented at the 5th Int. Conf. Ubiquitous Computing, Seattle, WA, 2003.
[34] C. R. Wren and S. G. Rao, “Self-Configuring Lightweight Sensor Networks for Ubiquitous Computing,” Mitsubishi Electric Research Laboratories, Cambridge, MA, Tech. Report TR2003-24, 2003.
[35] M. Langheinrich, V. Coroama, J. Bohn, and M. Rohs, *As We May Live—Real World Implications of Ubiquitous Computing*. ETH Zurich: Swiss Federal Inst. Technol., 2002, vol. 2003.
[36] P. Danielson, “Video surveillance for the rest of us: Proliferation, privacy, and ethics education,” in *Proc. IEEE 2002 Int. Symp. Technology and Society: Social Implications of Information and Communication Technology*, J. Herkert, Ed., 2002, pp. 162–167.
[37] S. J. McKenna, F. Marquis-Faulkes, P. Gregor, and A. F. Newell, “Scenario-based drama as a tool for investigating user requirements with application to home monitoring for elderly people,” presented at the HCI Int., Crete, Greece, 2003.

Alex Mihailidis received the B.A.Sc. degree in mechanical engineering and the M.A.Sc. degree in biomedical engineering from the University of Toronto, Toronto, ON, Canada, in 1996 and 1998, respectively, and the Ph.D. degree in biomedical engineering from the University of Strathclyde, Glasgow, U.K., in 2002

He is an Assistant Professor in the Department of Occupational Therapy and the Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada. His research focuses on the development of intelligent assistive technology and environments for older adults with cognitive and mobility impairments.

Brent Carmichael received the B.A.Sc. degree in mechanical engineering from the University of British Columbia, Vancouver, Canada, in 2003. He is working toward the M.A.Sc. degree in biomedical engineering at the Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada.

His research interests include computer vision, biomechanics, ergonomics, and assistive devices design.

Jennifer N. Boger received the B.A.Sc. degree in biological engineering from the University of Guelph, Guelph, Canada, in 2002. She is working toward the M.A.Sc. degree in biomedical engineering at the University of Toronto, Toronto, Canada.

Her thesis project is on the application of artificial intelligence based on Markov decision processes to guide people with dementia through activities of daily living.