Fifth Int'l Conference on Ubiquitous Computing (UbiComp'03), 2nd Int'l Workshop on Ubiquitous Computing for Pervasive Healthcare Applications, Seattle, WA, Oct 12-15, 2003.

Toward the Automatic Assessment of Behavioral Disturbances of Dementia

S. J. Allin[1], A. Bharucha[2], J. Zimmerman[1], D. Wilson[1], M. J. Roberson[1], S. Stevens[1], H. Wactlar[1] and C.G. Atkeson[1]

[1]Carnegie Mellon University, Pittsburgh PA
[2]Western Psychiatric Institute and Clinic, University of Pittsburgh, Pittsburgh PA

1 An Example of Pervasive Computing in Healthcare

Advances in healthcare have dramatically increased life expectancy in the United States over the course of the last century. The challenges of caring for a graying population are exemplified by Alzheimer‘s disease (AD), the most common cause of dementia among persons age 65 and older. The current annual direct and indirect cost of 100 billion dollars devoted to dementia care will increase as the number of Americans suffering from AD nearly triples to 14 million by the year 2050 [6].

Approximately 90% of individuals with dementia exhibit one or more behavioral disturbances over the course of the illness [3]. These behaviors can be broadly classified as physically aggressive, physically non-aggressive (e.g., wandering), verbally aggressive, and verbally non-aggressive (i.e., repetitive vocalizations). These symptoms not only reduce the quality of life of the individual with dementia and his/her caregiver, but also lead to premature institutionalization, inappropriate and excessive physical and chemical restraints and attendant medical consequences such as falls.

The psychosocial care of the nursing home resident is affected by inadequacies in staffing and training that are ubiquitous. As a result, clinical data often consists of brief observations of residents over relatively short periods of time, filtered through the lens of an overburdened staff member. In the absence of objective, reliable assessment and outcomes measurement methodologies, effectiveness of behavioral and pharmacological interventions cannot be determined.

Pervasive technology holds the promise of developing objective, real-time, continuous assessment and outcomes measurement methodologies that were previously unfeasible. Such technologies can contribute greatly to a deeper understanding of the activity and behavior patterns of individual residents, and the physical, environmental and psychosocial correlates of these patterns. These kinds of applications accord very well with research goals articulated by the International Psychogeriatric Association in 2002 [1].

2 Current Work

In February of 2002, researchers at Carnegie Mellon University began a collaboration with Dr. Ash Bharucha, a geriatric psychiatrist from the University of

Pittsburgh‘s Western Psychiatric Institute and Clinic, to assess the utility of pervasive technology in a dementia unit for the elderly. The group initially instrumented environments with cameras and microphones. This choice of sensors enables caregivers to see and hear recordings of episodes as directly as possible as they view statistical summaries. The group is also considering instrumenting wheelchairs, walkers, canes, and shoes to help identify, locate, and monitor patients. Problems are expected with wearable sensors that are not hidden from the patients or that require their cooperation, however. Residents of a dementia unit sometimes repeatedly undress and will forget to wear a badge. In addition, sensors requiring significant alterations to an individual‘s environment or routine may be less acceptable, especially among a population that is particularly sensitive to even small changes [4].

In the initial study, four cameras and microphones were mounted in the dementia unit of a local nursing home. The cameras recorded activity in public areas of the facility for eight hours a day over the course of one week. Four viewpoints were selected: one in the dining room, one in the main hallway adjacent to rooms, and the last two in the television room. Resulting videos were annotated by a group of undergraduate students under the supervision of Dr. Bharucha. Together, they scanned the data for visible or audible indicators of behavioral symptoms. These indicators were based upon those listed in the Cohen Mansfield Agitation Inventory as well as the Pittsburgh Agitation Scale. Both have established reliability and validity in the nursing home setting [1].

Based on the group‘s observation of videos, several initial goals for automation were identified. These include the quantitative study of:

1. **Interpersonal interactions**, such as when a patients talks with, is touched by or touches another.
2. **Changes in activity**, such as when a typically active patient fails to appear in group activities, or at meals.
3. **Wandering**, such as when a patient meanders into anothers room, appears lost, or makes an effort to escape.
4. **Physically and verbally aggressive actions**, such as when a patient talks loudly, hits, kicks, or curses.

Interactive designers led by John Zimmerman have been analyzing ways in which this kind of automation could ultimately serve caretakers. They conducted brainstorming sessions to identify discrete user groups and applications, and are following this up with an ethnographic study of caregivers in dementia wards. Prototype interfaces are now being constructed; these will be populated with data from the pilot study and evaluated by caregivers. An arm of research, then, has been aimed at discovering how proposed technologies can best meet the needs of caregivers.

In addition, three tiers of technical research have been formulated to address articulated goals. These are:

2.1 Level One: Use of Global Video and Audio Statistics

Image understanding is complicated, particularly when it is based on images taken outside of laboratory conditions. Even the best face detectors, for example, may fail to locate faces in high quality static images 25% of the time [7]. Locating a human being in a sequence of images can be equally complicated, particularly if the human in question is immobile. This is a common circumstance in nursing homes, where individuals may sit in common spaces for long periods. Separating a single voice from an audio stream is equally difficult, particularly when the audio is confounded by ambient noise from air conditioners and televisions.

In order to handle these kinds of technical challenges, statistics may be computed over an entire video or an aggregate audio stream, without ever making an effort to segment images into objects, or separate the audio into individual speakers. An example of research that takes this approach is found in the work of Zhang et. al. [10]. Zhang segmented 'related activities' in the video from the pilot study by computing simple, global statistics over short sub-sequences of video.

In a similar spirit, Allin and Wren used projection patterns to group similar sections of the video from the pilot study. A projection pattern is a normalized summation of the number of foreground pixels in a picture. This summation is computed across rows and columns of an image and stacked over time to form a sort of moving texture, or 'frieze' [8]. In their experiments, Allin and Wren used projection patterns to distinguish individuals from groups, patients from staff, and walkers from those engaged in conversation or standing still. A classification test involving 870 patterns, each two seconds in length, correctly distinguished walkers from others 84% of the time and individuals from groups 70% of the time.

Algorithms that operate on global image or audio statistics may be best suited to the detection of instances of **physically and verbally aggressive actions**.

2.2 Level Two: Use of Coarse Statistics Pertaining to an Individual

At level two, an effort is made to incorporate information as to who is in a given image or audio stream, where and at what time. Few assumptions are made about the physical configuration of individuals being tracked or the semantic content of their speech.

A human tracker has been developed which takes into account color of individuals and assumes smoothness in human motion. More specifically, it combines attributes of the mean-shift algorithm in color space [5] with a Kalman smoother in physical space. Mean shift tracking has been shown to be tolerant to shape changes in a target (as might occur when the target sits down) and it can recover when a subject turns away from the camera or becomes partially obscured. By compiling aggregate tracks for an individual over a whole day, a picture of typical movement patterns may be built with tools such as Hidden Markov Models. Ideally, such models can be used to detect anomalies.

As implemented, the tracker works sufficiently well to initialize a color target once motion has been detected, and to follow it. Yet errors still, inevitably, are made. The tracker occasionally comes to reside on objects in the background or becomes lost as people pass one another. Because of this, a part of our research seeks to determine what level of ' accuracy' is sufficient for our task. Can a tracker that falls off an individual 80% of the time yield reasonable estimates about his or her ability to ambulate? Can it be used to determine when he or she is trying to escape or in another persons room?

Algorithms which operate on image and audio statistics at this level of granularity may be best suited to detect instances of **wandering** and **changes in activity**.

2.3 Level Three: Use of Fine Grain Measurements and Statistics Pertaining to an Individual

Much research devoted to recognizing an individual's activities makes use of models built atop on prior knowledge of typifying kinematics and dynamics [9]. A good deal more relies on models built atop prior knowledge of token actions and sequences of these tokens [2]. Such representations are powerful, yet the models are complicated, difficult to initialize automatically and unstable, especially when scenes are cluttered and reflect the activity of groups.

At level three, researchers are seeking to make use of kinematic and probabilistic models to recognize detailed activities, actions, and events like falls. In addition, they seek to use models of affect and emotion that depend upon very high resolution images of faces, and high quality recordings of voices. In our next pilot study, plans have been made to utilize high resolution still cameras at dinner tables and in doorways. The hope is to populate acquired data with images that may reveal subtle and semantically relevant attributes of faces and hands. Finally, color based segmentations are being used to segment heads from torsos, and torsos from arms and legs. Activities such as falls or hits may ultimately be encoded based on relationships between these various atomic parts; these relationships can potentially be used to detect instances of **physically aggressive actions**, **interpersonal interaction** and **changes in activity**.

3 Expectations for the Workshop

The UBICOMP workshop provides the first author an opportunity to meet other researchers who are working with similar technologies and on similar issues. Moreover, it offers a valuable opportunity to achieve consensus among researchers as to problem areas in need of further research.

How other researchers determine accuracy in pervasive systems and relate it to overarching system goals is of interest. It is also of interest to know how others take into account the privacy of patients and the possibility that systems may contribute to paranoia and insecurity. We recognize that our sensor choice represents one of many potential configurations, each with merits and drawbacks.

4 Research Activities, Including Biographies

Sonya Allin is an NSF Graduate Fellow pursuing a Ph.D. in the Human Computer Interaction Institute of CMU. Her research is in the application of image understanding to activity recognition among the elderly. She also helped to assess force feedback devices‘ potential to physically rehabilitate stroke survivors.

Ashok Bharucha is a geriatric psychiatrist and Assistant Professor of Psychiatry at the Western Psychiatric Institute and Clinic. He specializes in nursing home psychiatry with a research emphasis on the behavioral symptoms of dementia.

John Zimmerman is an Assistant Professor at CMU with a joint appointment in the Human-Computer Interaction Institute and The School of Design.

Daniel Wilson is a third year Ph.D. student in the Robotics Institute at CMU who researches the ability of simple sensors to predict identity and activity of home occupants.

Matthew Johnson Roberson is in his third year of an undergraduate degree in Computer Science at CMU.

Scott M. Stevens is Senior Systems Scientist in the School of Computer Science, CMU. He is a Co-PI of the Informedia Digital Library Project.

Howard Wactlar is Vice Provost for Research Computing, Associate Dean, and Principal Research Scientist in the School of Computer Science, CMU.

Chris Atkeson is a CMU Associate Professor in Robotics and HCI. His previous projects include Georgia Tech's Aware Home and Classroom 2000.

References

1. Behavioral and psychological symptoms of demention educational pask. *International Psychogeriatric Association*, 2002.
2. M. Brand and I. Essa. Causal analysis for visual gesture understanding. *AAAI Fall Symposium on Computational Models for Integrating Language and Vision*, 1995.
3. J. A. Brody. An epidemiologist views senile dementia: facts and figures. *American Journal of Epidemiology*, 113:155–162, 1982.
4. L. Burgio, K. Scilley, et al. Temporal patterns of disruptive vocalization in elderly nursing home residents. *International Journal of Geriatric Psychiatry*, 16:378-386, 2001.
5. D. Commaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. *CVPR*, 2142-2145, 2000.
6. J. W. Ernst, R. L. Hay and C. Fenn. Cognitive function and the costs of alzheimer's disease: an exploratory study. *Arch Neurol*, 54(6):687-693, 1997.
7. R. Jin and A. Hauptmann. Learning to identify video shots with people based on face detection. *IEEE Int Conf on Multimedia and Expo*, 2003.
8. Y. Liu, R. Collins, and Y. Tsin. Gait sequence analysis using frieze patterns. *ECCV*, Copenhagen:657–671, 2002.
9. H. Sidenbladh and M. Black. Learning the statistics of people in images and video. *International Journal of Computer Vision*, 54(1-3):183-2, 2003.
10. H. Zhang and J. Shi. Finding (un)usual events in video. *Carnegie Mellon University*, CMU-RI-TR-03-0, 2003.