Automatic Collection, Analysis, Access, and Archiving of Psycho/Social Behavior by Individuals and Groups

Scott Stevens[1], Ashok Bharucha[2], Michael Christel[3], Alexander. Hauptmann[3,] Howard Wactlar[3], Datong Chen[3]

[1]Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213, sms@cs.cmu.edu

[2]Western Psychiatric Institute and Clinic, University of Pittsburgh, Pittsburgh, PA 15213

[3]Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213

[4]Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213

Nearly 2.5 million Americans currently reside in nursing homes and assisted living facilities in the United States, accounting for approximately 5% of persons 65 years and older.[1]The aging of the "Baby Boomer" generation is expected to lead to an exponential growth in the need for some form of long-term care (LTC) for this segment of the population within the next 25 years. In light of these sobering demographic shifts, there is an urgency to address the profound concerns that exist about the quality-of-care (QoC) and quality-of-life (QoL) of this frailest segment of our population.

CareMedia [CMUInf04]is an interdisciplinary, collaborative effort that to date has captured more than 13,000 hours of video and audio recordings of life in the shared spaces of a nursing home dementia unit, by using 23 ceiling-mounted cameras, 24 hours a day for 25 days, ensuring an un-occluded view of every point in the recorded space (Fig. 1). Computer machine learning techniques are being applied to the resulting 25 Terabytes of data, automatically processing the record for efficient use by analytical observers (e.g., social and behavioral scientists, geriatricians, and healthcare policymakers) to monitor and understand residents' well-being, and enhance their QoL. This truly interdisciplinary effort bridges the psychological, social and behavioral sciences, and clinical medicine with multiple engineering and computer science disciplines to establish a clinical evidence base to guide rational therapeutics, an elusive goal ardently articulated by the Institute of Medicine.

The record is currently being processed with data reduction and extraction technologies that recognize faces and speech, track moving individuals, and identify a range of human activities and social interactions. The automated audiovisual record provides a level of completeness not feasible with human observers, and facilitates, for the first time, large-scale longitudinal research that captures the full range of the biopsychosocial context in elder care settings, thus refining ecological, ethnographic, qualitative and quantitative aspects of research methodology (Fig 2). Ultimately, these methods will automatically recognize, classify, and quantify both characteristic (baseline) and aberrant affect,

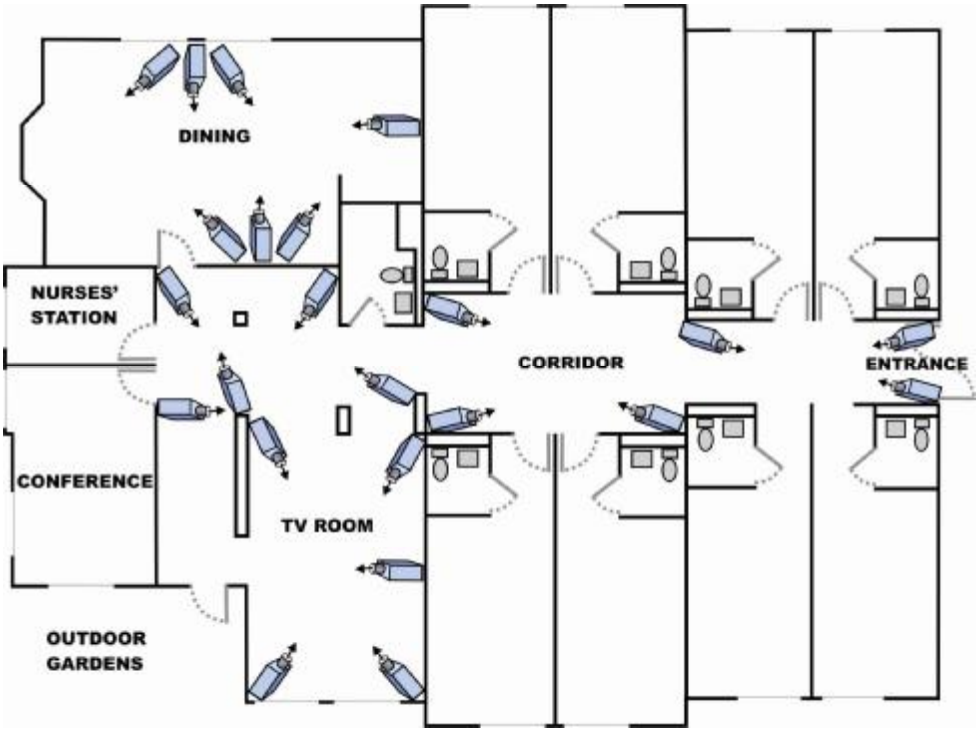

Figure 1. Facility Architecture and Camera/Microphone Instrumentation

behavior, motor function, and activity patterns continuously over extended periods of time (days to weeks to months), thereby advancing the state of biobehavioral research and care by developing tools that:

- Augment qualitative observations with quantitative dimensions, thus transforming largely subjective assessments into more measured, objective ones.
- Enrich discrete, human observations with a machine-captured, continuous longitudinal record complete with automated analysis.
- Detect and annotate the possible antecedents and consequences of salient events.
- Detect and trace the trajectory of subtle changes in patient functioning.
- Refine and expand existing methodologies for coding affect, behavior, and social interactions.

- Address the subtle recruitment challenges characteristic of geriatric clinical studies (small sample sizes) by increasing the accuracy and precision of measurements, thus increasing statistical power.

Such enabling tools will overcome the methodological challenges and limitations of present clinical research, particularly as it relates to behavior, and especially for a population unable to provide accurate self-reporting.

Viewing thousands of hours of video is a daunting challenge that is clearly prohibitive. We will address this task by providing a continuous but summarized, automatically processed audiovisual record that will assist clinical decision-making, enabling earlier interventions, and providing new measures of their effectiveness. The current robustness of machine vision and audio cognition is subject to environmental conditions and heterogeneity of the target population. Machine learning is effective when the set of classifications and outcomes is finite and known. The restricted residential setting provided by nursing homes is hence an ideal focus environment as a starting point for this research, ultimately moving towards applications in the home setting. The automated vision cognition tasks range from patient identification and tracking through measurement of residents' mobility, eating behavior and social interactions, to classification and quantification of gait, abnormal movements and facial expressions. Audio understanding tasks encompass classification of environmental and human sounds through identification of speakers to recognition of individual and group speech in noisy surroundings. The computer science research is focusing on the demands and challenges of machine understanding for applications related to the elder care setting. Figure 2 illustrates graphically the information flow and processing in the CareMedia system.

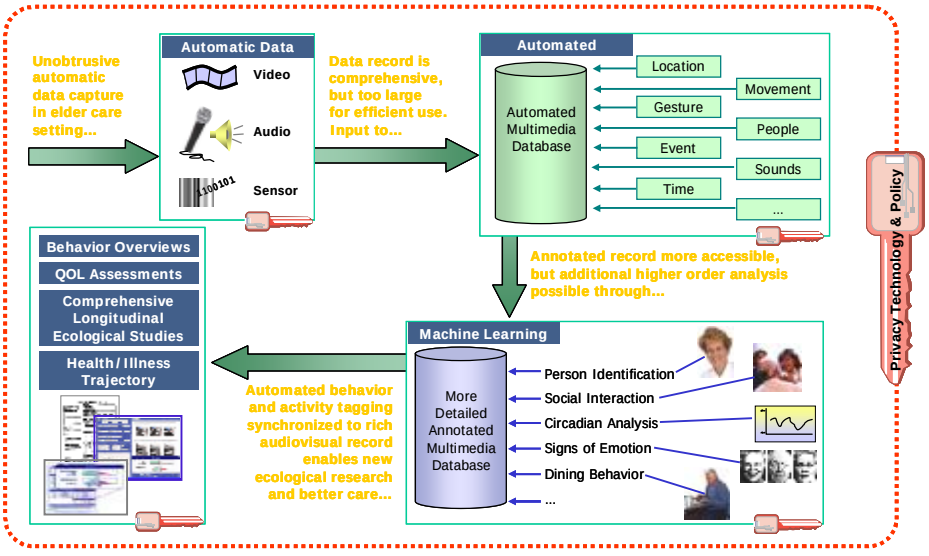

Figure 2: Conceptual Overview of CareMedia

CareMedia builds on over 12 years of CMU Informedia digital Video Library (IDVL) research to enable useful access to such voluminous data sets [Stevens96]. The IDVL

interface has been modified for CarMedia applications in several ways. CareMedia requires training data comprised of human coding of day-to-day activities and interactions in order to generate 'ground truth' to permit quantitative analysis of the automated processing. Four graduate students in the social and behavior sciences are coding frame-by-frame audiovisual recordings from selected cameras and specified time intervals (Fig. 3).

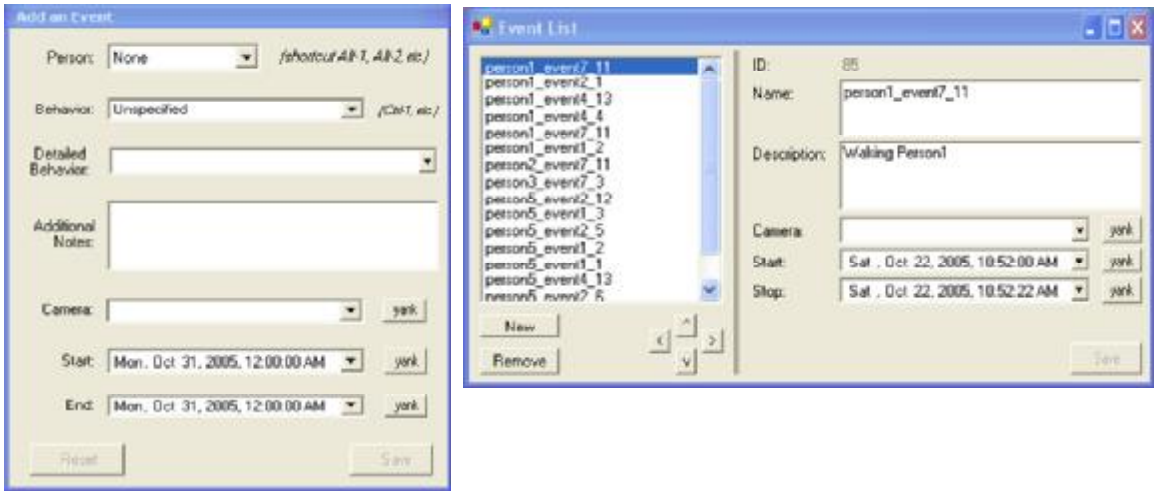

Figure 3. CareMedia's Manual Coding Interface

The first living space being examined is the dining room. Each coder is assigned to one dining table with up to a maximum of 5 residents per table for 25 lunch intervals (total 2.5 hours a day from 10:30 AM – 1 PM) captured during our study period. This manual coding occurs in two passes, the first caputring movement and the second more complex human behaviors. In a previous CareMedia study, our cameras captured 6 un-witnessed elopements from this locked dementia unit, usually behind unsuspecting visitors and staff to the unit [Bharucha06]. Since automatic detection of exit-seeking behavior is of great safety interest to caregivers, entrances and exits were examined next. Coders noted every entrance and exit and the identify of the person involved for four sets of doors for 24 hours: the main entrance to the dementia unit, entrance to the dining and living rooms, and exit to the garden.

An event for our purposes is defined as any human activity or occurrence that has important implications for clinical assessment and management, e.g. a fall. Unlike the frame-by-frame human processing of the data, automatic event processing of necessity generates many false positive and false negative results. For example, a person walking through the hallway, a moving shadow, or even a simple flicker of a light may register as an event if the specified event is entrances into the dining room (Fig 4).

The inordinate number of events generated by even brief periods of observation would overwhelm any user of these data sets. In order to facilitate meaningful processing and analysis of these data sets efficiently, 'filters' are developed to assist end users in this process. For example, identifying faces is an obvious and useful metric to eliminate many of the false positive results. Figure 5 illustrates image filtering based, in this case,

on faces narrowing the results to a more limited, applicable data set. By moving the slider to the left (reducing the processing system's confidence level), a greater number of possible dining room entrance events can be generated, increasing recall at the expense of precision.

This demonstration highlights the features of the CareMedia user interface as well as simple and complex filtering and search techniques applied to a subset of the full data set. We present a preliminary analysis of precision and recall in CareMedia, the multiple perspective from which the data can be accessed and viewed, and discuss the advantages and limitations of applied machine learning to this unique corpus.

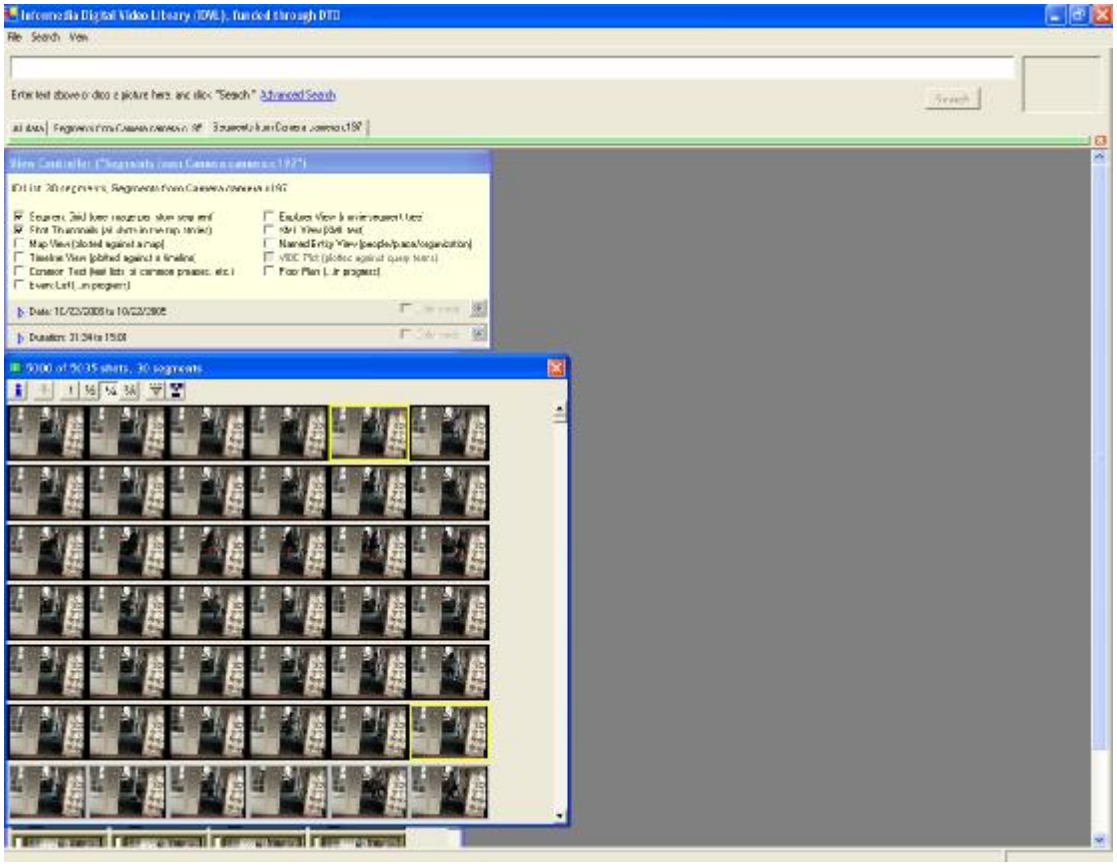

Figure 4. Dining Room Entrance 'events.'

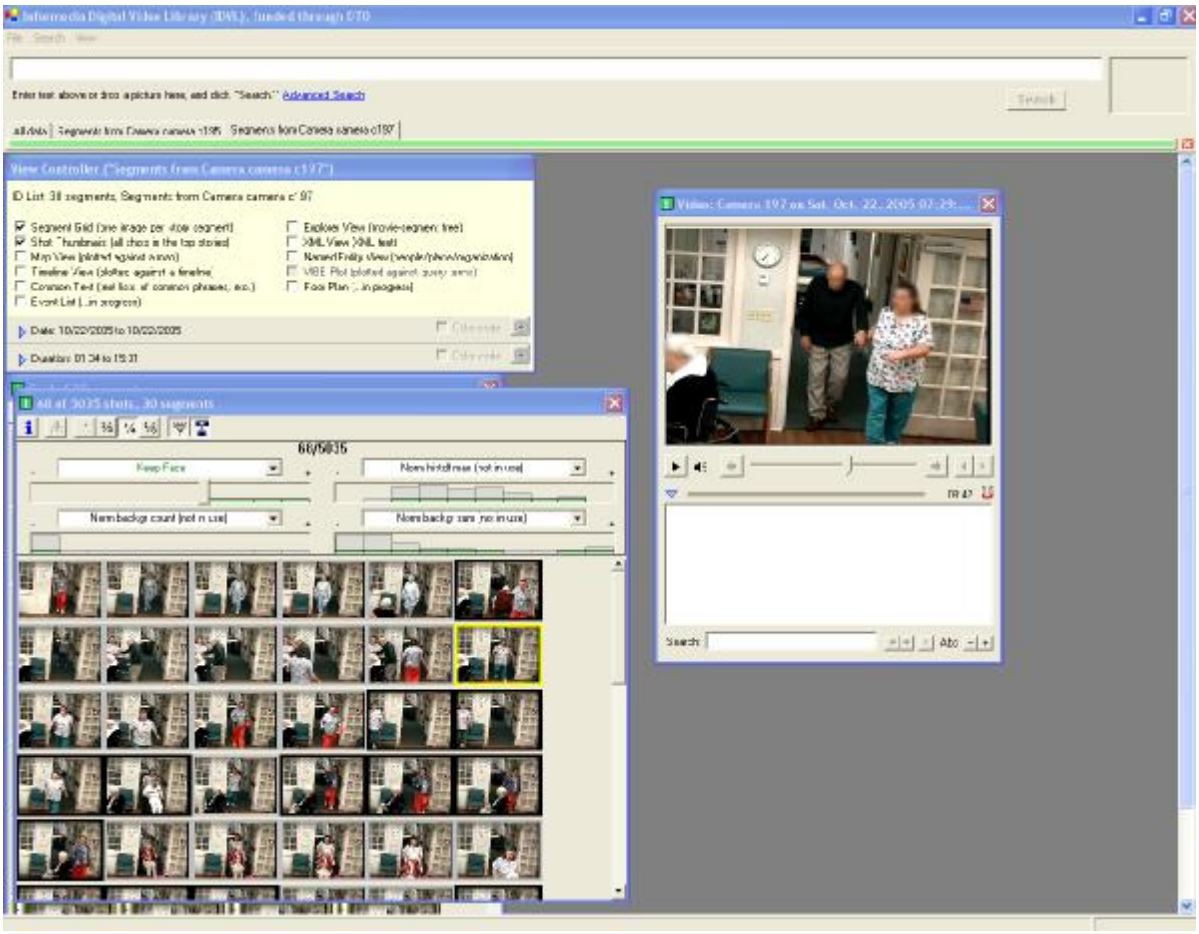

Figure 5. Application of the Face Filter to the Dining Room Entrance

REFERENCES

[Bharucha06] Bharucha AJ, Wactlar H, Stevens S, Chen D, Atkeson C, Dew MA, Pollock BG, Reynolds CF. *Preliminary Evaluation of Automatically Analyzed Video and Sensor Data for Geriatric Care,* Annual Meeting of the American Medical Directors Association, Dallas, TX, March 17, 2006.

[Borson00] Moak G, Borson S. Mental Health Services in Long-Term care. *Am J Geriatr Psychiatry* 2000;8:96-101.

[CMUInf04] Informedia Carnegie Mellon University Informedia Research Group Web Site, 2004. http://www.informedia.cs.cmu.edu

[Stevens96] Stevens,S.M., & Wactlar, H., インフォメディア・ディジタル・ビデオ・ライブラリ・プロジェクトにおける応用の研究

("New Models for Human–Computer Interaction in Digital Video Libraries,") In Ichikawa & Yoshitaka, Ed. *Multimedia Computing: Supporting Human Activities,* Special edition of *bit,* Kyoritsu Publishing Co., Tokyo, Japan, August 1996, 154-162