

**CareMedia:  
Automated Video and Sensor  
Analysis for Geriatric Care**

**NSF Cooperative Agreement No. IIS-0205219  
Annual Progress Report  
March 2004**

**Carnegie Mellon University  
School of Computer Science  
Pittsburgh, PA 15213-3890**

**Principal Investigator:** **Howard Wactlar**  
School of Computer Science

**Co-PI's:**

<b>Alexander Hauptmann</b> Computer Science Dept.	<b>Takeo Kanade</b> Robotics Institute
<b>Scott Stevens</b> Computer Science Dept.	<b>Ashok Bharucha</b> Western Psychiatric Institute

## 1 Project Participants

*What people have worked on the project?*

Participant's Name	Project Role	>160 Hours?
Howard D. Wactlar	Principal Investigator	Yes
Takeo Kanade	Co-Principal Investigator	Yes
Ashok Bharucha	Co-Principal Investigator	Yes
Scott M. Stevens	Co-Principal Investigator	Yes
Alexander G. Hauptmann	Co-Principal Investigator	Yes
Sonya Allin	Graduate Student	Yes
Christopher Atkeson	Project Scientist	Yes

*Are any other persons involved with the project?*

None.

*What other organizations have been involved as partners?*

University of Pittsburgh Medical Center (UPMC).

Western Psychiatric Institute and Clinic (WPIC).

*Have you had other collaborators or contacts?*

Longwood at Oakmont Retirement Resort, Dementia Care Facility

## 2 Project Activities and Findings

### Overview

This research will create a continuously recorded, digital history of patients' activities in a nursing home by capturing all that is heard, seen and experienced. Our research challenge is to transform a voluminous amount of captured video, audio and sensor data into a meaningful information resource that enables more complete and accurate assessment, diagnosis, treatment, and evaluation of behavioral problems for the elderly. The proposed work in (i) information extraction and retrieval, (ii) behavior recognition, analysis and summarization, and (iii) secure, efficient visual information access will let geriatric care specialists obtain greater insights into problems, effectiveness of treatments, and environmental and social influences on patient behavior. Prototype systems for activity monitoring and behavior analysis will be deployed at local area nursing homes and dementia wards to be utilized by medical professionals in trials conducted by our project partners from the University of Pittsburgh Medical Center (UPMC), Western Psychiatric Institute and Clinic (WPIC).

## 2.1 Activities

We have continued to experiment on preliminary “surveillance-type” video, gathered at a nursing care facility, Longwood at Oakmont’s Dementia Care Facility, under the supervision of our psychiatric colleague, Dr. Ashok Bharucha. We exploit this set of data extensively, and have designed and prototyped an interface into the massive video record, allowing clinicians and others ready access to items of interest. Additionally, we simulated a nursing home environment at Carnegie Mellon, duplicating lighting conditions and occlusion problems we anticipate in our future large-scale deployment. Using this captured video data, we developed a technique to track multiple individuals across multiple camera views.

### 2.1.1 Toward the Automatic Assessment of Behavioral Disturbances of Dementia

The psychosocial care of the nursing home resident is affected by inadequacies in staffing and training that are ubiquitous. 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.

The CareMedia project uses global video and audio statistics computed over a video stream to group together similar sections of video, mean shift tracking to compile aggregate tracks of an individual’s motion in 2D space, and fine grain kinematic and probabilistic models to recognize detailed activities, actions, and events like falls.

### 2.1.2 Detecting Unusual Activity in Video

In surveillance applications, unusual events are those that should be reported for further examination. Unusual events are by their nature rare, difficult to describe, hard to predict and can be subtle. However, given a large number of observations it is relatively easy to verify if they are indeed unusual.

We have developed a method to utilize the *hard to describe* but *easy to verify* property of unusual events without building explicit models of normal events. One can compare each event with all other events observed to determine how many *similar* events exist. If an event is normal, there should be many similar events in this large data set. If there are no similar events we consider this unusual: although the event is unknown, it is different from the others. Thus, detecting unusual events in a large data set does not require modeling normal events, but rather the ability to compare two events and measure their similarity.

This unsupervised technique exploits simple features. No complex activity models and no supervised feature selections are used. We divide the video into equal length segments and classify the extracted features into prototypes, from which a prototype–segment co-occurrence matrix is computed. Motivated by a similar problem in document keyword analysis, we seek a correspondence relationship between prototypes and video segments which satisfies the transitive closure constraint. We have demonstrated that an important sub-family of correspondence functions can be reduced to co-embedding prototypes and segments to N-D Euclidean space, and proved that an efficient, globally optimal algorithm exists for the co-embedding problem. Experiments on various real-life videos have validated our approach.

### **2.1.3 Multiple Camera Angles**

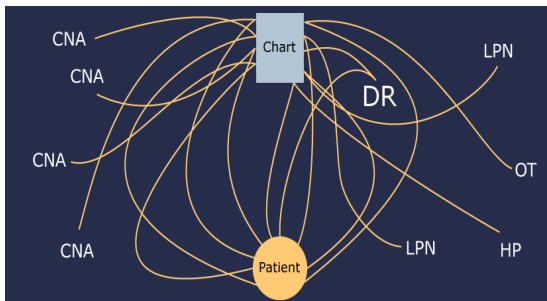
Using a simulated nursing home video, we have developed a robust technique for tracking multiple people across a camera network. Person location and tracking provide crucial information for analyzing patient walking activities and social interactions, such as how often and how fast a patient walked every day or every week. Does a patient need to use a wheelchair? How many social interactions has a patient had today? To answer these questions, a tracking system should provide a walking (or wheelchair) trace corresponding to the geometric information of the nursing home. The tracking information needs to be recovered in 3D space so that it can be mapped onto the map of the nursing home. A primary difficulty for tracking people in a nursing home comes from the occlusions between multiple people who share the monitored public areas in the space.

To address the multiple people occlusion problem, we developed a tracking technique based on 3D walking prediction and Monte Carlo search using particle filtering. This technique uses one or more cameras to cover the target area. The location of a person in 3D space is tracked by integrating tracking confidence from all the cameras. The position of a reference point of a person is first predicted in a common 3D world coordination using particle filters. The predicted 3D position is then mapped onto multiple camera images. Tracking confidences are finally computed at mapped positions in all the camera images and integrated together for robust tracking. The advantage of this technique is that it can optimize the tracking results through an unknown number of cameras. Traditional stereo algorithm based tracking directly predicts an object's motion in 2D camera images and then makes a hard tracking decision for each of the cameras. The 3D information is computed using the tracking results of 2 or 3 cameras based on spatial geometric relations. The hard decision step may yield tracking errors that will be further amplified by the following 3D recovery step. Although multiple hypotheses can be made instead of hard decisions, the computation of the following 3D recovery step is exponential to the number of hypotheses. Furthermore, it is not easy to reduce the tracking error in 3D space by using more than 3 cameras since the 3D recovery step is also a hard decision maker, which means the error has a complex distribution. Our tracking technique

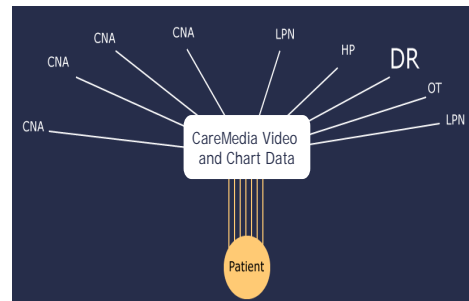
can avoid making hard decisions in both the camera level and 3D recovery level. There is no special requirement for the number of cameras in which the target person is tractable. Physical constraints (for example, the height of the person or special walking patterns) can be easily integrated into the proposed technique, which is another advantage. Experiments show that the average tracking error reduction ratio is about 58.5% by using this technique with three cameras compared to a single camera.

#### 2.1.4 Interface Development

In order to better understand the value a system like CareMedia can bring to both caregivers and patients, we conducted a series of shadowing exercises. We observed caregivers in a dementia ward in order to better understand their information flow and information needs. We also shadowed geriatric psychiatrists as they conducted rounds. The information we collected will help us design a system that provides relevant data at appropriate times without interrupting the current work procedures. Figure 1 shows the current, somewhat chaotic flow of chart-based patient information, reflecting the observed fact that individual caregivers use patient chart to access info relevant to their job, ignoring rich data collected by other types of staff. Figure 2 shows a more optimal, CareMedia based information flow, which provides a collective visual record for each patient. Caregivers can both complete their tasks and benefit from colleagues' input.



*Figure 1: Each caregiver accesses a patient chart for specific information, ignoring rich data collected by other types of staff.*



*Figure 2: CareMedia offers access to a collective visual record for each patient. Caregivers benefit from colleagues' input.*

After the shadowing exercises we developed two prototype interfaces. The first (Figure 3) is designed for caregivers in the ward. It is meant to provide up to the minute information on patients and support the nurses and nurses aids who spend considerably more time in the ward than doctors. The second interface focuses on summarizing a day in the life of a single patient. It visualizes all of their social and aggressive interactions and can be used to view video clips of these interactions for assessment purposes. We are currently meeting with doctors and nurse practitioners to assess if this design both 1) meets their informational and 2) fits within the time constraints of their current work flow.

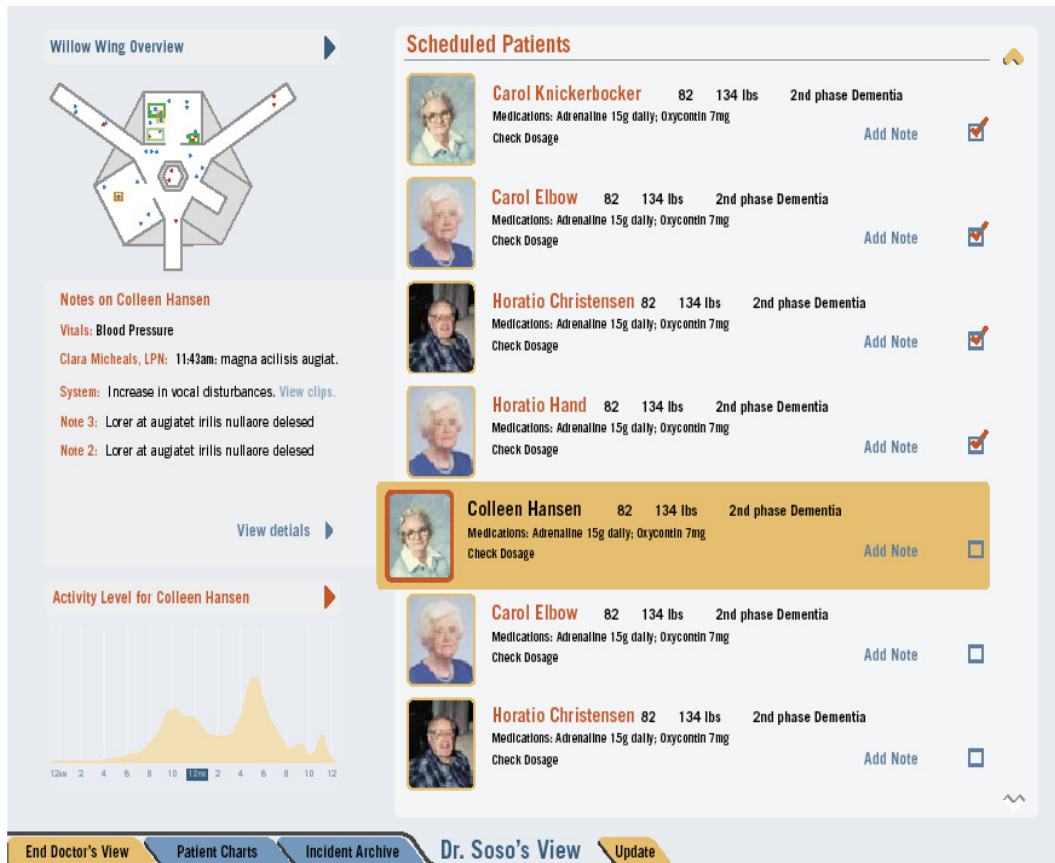


Figure 3: Prototype CareMedia interface developed for caregiver assessment.

### 2.1.5 Automated Analyses of Nursing Home Observations

Our research analyzes nursing home video information by automatically tracking people, assisting in efficiently labeling individuals, and characterizing selected activities and actions. Special emphasis is given to detecting eating activity in the dining hall and to personal hygiene. Through this work, the video record is transformed into an information asset that can provide geriatric care specialists with greater insights and evaluation of behavioral problems for the elderly. Evaluations of the effectiveness of analyzing such a large video record illustrate the feasibility of our approach.

### 2.1.6 Combining Motion Segmentation with Tracking for Activity Analysis

We explored a novel motion feature as the appropriate basis for classifying or describing a number of fine motor human activities. Our approach not only estimates motion directions and magnitudes in different image regions, but also provides accurate segmentation of moving regions. Through a combination of motion segmentation and region tracking techniques, while filtering for temporal consistency, we achieve a balance between accuracy and reliability of motion feature extraction. To identify specific activities, we characterize the dominant directions of relative motions. Experimental results show that this approach to

motion feature analysis could be successful in assisting caregivers at a nursing home in assessments of patient's activity levels over time.

### **2.1.7 Dining Activity Analysis Using Hidden Markov Models**

We developed an algorithm for dining activity analysis in a nursing home. Based on several features, including motion vectors and distance between moving regions in the subspace of an individual person, a hidden Markov model is proposed to characterize different stages in dining activities with certain temporal order. Using HMM model, we are able to identify the start (and ending) of individual dining events with high accuracy and low false positive rate. This approach could be successful in assisting caregivers in assessments of resident's activity levels over time.

## **2.2 Findings**

Allin, S.J., Bharucha, A., Zimmerman, J., Wilson, D., Roberson, M.J., Stevens, S., Wactlar, H., and Atkeson, C.G., "Toward the Automatic Assessment of Behavioral Disturbances of Dementia." Proceedings of the Fifth International Conference on Ubiquitous Computing (UbiComp'03) and the Second International Workshop on Ubiquitous Computing for Pervasive Healthcare Applications, Seattle, WA, Oct. 12-15, 2003.

Abstract: In the absence of objective, reliable assessment and outcomes measurement methodologies in a nursing home, 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.

Bharucha, A., Allin, S. and Stevens, S., "CareMedia: Towards Automated Behavior Analysis in the Nursing Home Setting," in The International Psychogeriatric Association Eleventh International Conference, Aug. 17-22, 2003.

Several hours of surveillance-type video were captured in a nursing home. The task of data reduction and extraction of high-level activity information was approached through both automated and manual techniques. For the manual encoding, 4 undergraduate students were trained by a geriatric psychiatrist to code the data frame-by-frame. A computer interface allowed coders to annotate behaviors of interest, as well as physical pose and ambulatory status. Behaviors of interest were identified with the Cohen-Mansfield Agitation Inventory and grouped into 4 sub-categories:

physically aggressive, physically non-aggressive, verbally aggressive, and verbally non-aggressive. These manual encodings are currently forming the development of automated techniques at Carnegie Mellon University to extract information relevant to the detection of anomalous and disruptive physical activities. This includes automated tracking and extraction of navigational patterns.

Gao, J., Hauptmann, A.G., Barucha, A. and Wactlar, H.D., "Dining Activity Analysis Using Hidden Markov Models," accepted to The 17th International Conference on Pattern Recognition (ICPR'04), Cambridge, United Kingdom, Aug. 23-26, 2004.

Abstract: We describe an algorithm for dining activity analysis in a nursing home. Based on several features, including motion vectors and distance between moving regions in the subspace of an individual person, a hidden Markov model is proposed to characterize different stages in dining activities with certain temporal order. Using HMM model, we are able to identify the start (and ending) of individual dining events with high accuracy and low false positive rate. This approach could be successful in assisting caregivers in assessments of resident's activity levels over time.

Gao, J., Hauptmann, A.G. and Wactlar, H.D., "Combining Motion Segmentation with Tracking for Activity Analysis," submitted to The Sixth International Conference on Automatic Face and Gesture Recognition (FG'04), Seoul, Korea, May 17-19, 2004.

Abstract: We explore a novel motion feature as the appropriate basis for classifying or describing a number of fine motor human activities. Our approach not only estimates motion directions and magnitudes in different image regions, but also provides accurate segmentation of moving regions. Through a combination of motion segmentation and region tracking techniques, while filtering for temporal consistency, we achieve a balance between accuracy and reliability of motion feature extraction. To identify specific activities, we characterize the dominant directions of relative motions. Experimental results show that this approach to motion feature analysis could be successful in assisting caregivers at a nursing home in assessments of patient's activity levels over time.

Hauptmann, A.G., Gao, J., Yan, R., Qi, Y., Yang, J., and Wactlar, H.D., "Aiding Geriatric Patients and Caregivers through Automated Analysis of Nursing Home Observations," to be published in IEEE Pervasive Computing, April-June special issue: Pervasive Computing for Successful Aging.

Abstract: Through pervasive activity monitoring in a skilled nursing facility, a continuous audio and video record is captured.

Our CareMedia Project research analyzes this video information by automatically tracking people, assisting in efficiently labeling individuals, and characterizing selected activities and actions. Special emphasis is given to detecting eating activity in the dining hall and to personal hygiene. Through this work, the video record is transformed into an information asset that can provide geriatric care specialists with greater insights and evaluation of behavioral problems for the elderly. Evaluations of the effectiveness of analyzing such a large video record illustrate the feasibility of our approach.

Hauptmann, A.G., Jin, R. and Wactlar, H.D., "Data Analysis for a Multimedia Library, in Text and Speech-Triggered Information Access," Renals, S and Grefenstette, G. (eds)., Springer, Berlin, pp. 6-37, 2003.

Abstract: This book section describes the indexing, search and retrieval of various combinations of audio, video, text and image media and the automated content processing that enables it. The intent is to provide a framework for data analysis in multimedia digital libraries. The introduction briefly distinguishes the digital from traditional libraries and touches on the specific issues important to searching the content of multimedia libraries. The second section introduces the Informedia Digital Video Library as an example of a multimedia library, including a quick tour of the functionality. The next section discusses the processing of audio and image information, as it relates to a multimedia library. Section four illustrates the interplay between audio and video information using a video information retrieval experiment as an example. Section five discusses the exporting and sharing of metadata in a digital library using MPEG-7. Finally, section 6 provides one vision of a future digital library, where all personal memory can be recorded and accessed.

Jin, R., Hauptmann, A., Carbonell, J., Si, L., Liu, Y., "A New Boosting Algorithm Using Input Dependent Regularizer," 20th International Conference on Machine Learning (ICML'03), Washington, DC, August 21-24, 2003.

Abstract: AdaBoost has proved to be an effective method to improve the performance of base classifiers both theoretically and empirically. However, previous studies have shown that AdaBoost might suffer from the overfitting problem, especially for noisy data. In addition, most current work on boosting assumes that the combination weights are fixed constants and therefore does not take particular input patterns into consideration. In this paper, we present a new boosting algorithm, "WeightBoost", which tries to solve these two problems by introducing an input-dependent regularization factor to the combination weight. Similarly to AdaBoost, we derive a learning procedure for WeightBoost, which

is guaranteed to minimize training errors. Empirical studies on eight different UCI data sets and one text categorization data set show that WeightBoost almost always achieves a considerably better classification accuracy than AdaBoost. Furthermore, experiments on data with artificially controlled noise indicate that WeightBoost is more robust to noise than AdaBoost.

Wactlar, H., Bharucha, A., Stevens, S., Hauptmann, A., Christel, M., "A System of Video Information Capture, Indexing and Retrieval for Interpreting Human Activity," Third International Symposium on Image and Signal Processing and Analysis (ISPA'03), Invited paper for the Special Session on System Perspectives in Information Retrieval (to be published), September 18-20, 2003.

Abstract: This system creates a manageable information resource that enables more complete and accurate interpretation, assessment and diagnosis of human behavior in constrained physical spaces. Through activity and environmental monitoring, a continuous, voluminous audio and video record is captured. Through work in information extraction, behavior analysis and synthesis, this record is transformed into an information asset whose efficient, secure presentation empowers specialists with greater insights into problems, effectiveness of treatments, and determination of environmental and social influences. Application environments range from nursery schools to nursing homes. The foundation for this work, the Informedia Digital Video Library, has demonstrated the successful integration of speech, image, and natural language processing in automatically creating an indexed, searchable multimedia information resource for broadcast-quality video, upon which this system builds.

Yan, R., Hauptmann, A., "Automatically Labeling Data Using Using Multi-class Active Learning," Ninth International Conference on Computer Vision (ICCV'03), Nice, France, pp. 516-523, October 13-16, 2003.

Abstract: Labeling video data is an essential prerequisite for many vision applications that depend on training data, such as visual information retrieval, object recognition, and human activity modeling. However, manually creating labels is not only time-consuming but also subject to human errors, and eventually, becomes impossible for a very large amount of data (e.g. 24/7 surveillance video). To minimize the human effort in labeling, we propose a unified multi-class active learning approach for automatically labeling video data. The contributions of this paper include extending active learning from binary classes to multiple classes and evaluating several practical sample selection strategies. The experimental results show that the proposed approach works effectively even with a significantly reduced amount of labeled

data. The best sample selection strategy can achieve more than a 50% error reduction over random sample selection.

Yan, R., Hauptmann, A., "The Combination Limit of Video Retrieval," ACM Multimedia 2003 (MM'03), Berkeley, CA, November 2-8, 2003.

Abstract: Combining search results from multimedia sources is crucial for dealing with heterogeneous multimedia data, particularly in multimedia retrieval where a final ranked list of items of interest is returned sorted by confidence or relevance. However, relatively little attention has been given to combination functions, especially their upper bound performance limits. This paper presents a theoretical framework for studying upper bounds for two types of combination functions. A general upper bound and two approximations are proposed for monotonic combination functions. We also studied the upper bounds for linear combination functions using a global optimization technique. Our experimental results show that the choice of combination functions has a considerable influence to retrieval performance.

Yan R., Liu, Y., Jin, R., Hauptmann, A., "On Predicting Rare Class with SVM Ensemble in Scene Classification," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03), Hong Kong, China, April 6-10, 2003.

Abstract: Scene classification is an important technique to infer high-level semantic scene categories from low-level visual features. However, in the real world the positive data for many scenes may be rare, which degrades the performance of many classifiers. In this paper, we propose SVM ensembles to address the rare class problem. Various classifier combination strategies are investigated, including majority voting, sum rule, neural network gater and hierarchical SVMs. We also compare our method with two other common approaches for dealing with the rare class problem. Our experimental results show that hierarchical SVMs can achieve significantly better and more stable performance than other strategies, as well as high computational efficiency.

Zhong, H., Shi, J., and Visontai, M., "Detecting Unusual Events in Video," to appear in the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04), Washington, DC.

Abstract: We propose an algorithm for detecting and categorizing (un)usual human activity in a video which might be a few days long. The proposed approach is unsupervised, and uses the co-occurrence among large number of simple visual image features to define the activity categories, and to identify what are unusual events automatically. A video is divided into short segments (clips), and motion/color histogram is extracted for the foreground object for

each image frame. The image features are Vector Quantized into a smaller set of prototype features. A weighted graph is constructed by taking clips and prototype features as nodes, and the co-occurrence relationship between them as the graph edges. We compute an optimal graph embedding that maps the clips and prototype features in a common low dimensional space. This unified embedding ensures that all pair of co-occurring clip and feature are as close as possible. We define event categories by identify cluster of clips in this embedding space, and those isolated clusters are detected as unusual events. We can also classify a new video clips based on the embedding of its co-occurring features. We demonstrated this algorithm on several long surveillance video recorded at a nursing home.

### **3 Training and Development**

Graduate student support has enabled research into object and people tracking, and activity and behavior classification in video.

Clinical psychiatrist Dr. Ashok Bharucha has been introduced to computation techniques for capturing behavior in a clinical setting.

Undergraduate students have been exposed to techniques for video analysis during the analysis phase of our pilot study.

### **4 Outreach Activities**

Our work at a nursing home during the pilot study has exposed staff and patient families to the potential clinical value of computational technology.

### **5 Publications and Products**

Allin, S.J., Bharucha, A., Zimmerman, J., Wilson, D., Roberson, M.J., Stevens, S., Wactlar, H., and Atkeson, C.G., "Toward the Automatic Assessment of Behavioral Disturbances of Dementia." Proceedings of the Fifth International Conference on Ubiquitous Computing (UbiComp'03) and the Second International Workshop on Ubiquitous Computing for Pervasive Healthcare Applications, Seattle, WA, Oct. 12-15, 2003.

Bharucha, A., Allin, S. and Stevens, S., "CareMedia: Towards Automated Behavior Analysis in the Nursing Home Setting," in The International Psychogeriatric Association Eleventh International Conference, Aug. 17-22, 2003.

Gao, J., Hauptmann, A.G., Barucha, A. and Wactlar, H.D., "Dining Activity Analysis Using Hidden Markov Models," accepted to The 17th International Conference on Pattern Recognition (ICPR'04), Cambridge, United Kingdom, Aug. 23-26, 2004.

Gao, J., Hauptmann, A.G. and Wactlar, H.D., "Combining Motion Segmentation with Tracking for Activity Analysis," submitted to The Sixth International Conference on Automatic Face and Gesture Recognition (FG'04), Seoul, Korea, May 17-19, 2004.

Hauptmann, A.G., Gao, J., Yan, R., Qi, Y., Yang, J., and Wactlar, H.D., "Aiding Geriatric Patients and Caregivers through Automated Analysis of Nursing Home Observations," to be published in IEEE Pervasive Computing, April-June special issue: Pervasive Computing for Successful Aging.

Wactlar, H., Bharucha, A., Stevens, S., Hauptmann, A., Christel, M., "A System of Video Information Capture, Indexing and Retrieval for Interpreting Human Activity," Third International Symposium on Image and Signal Processing and Analysis (ISPA'03), Invited paper for the Special Session on System Perspectives in Information Retrieval, September 18-20, 2003.

Zhong, H., Shi, J., and Visontai, M., "Detecting Unusual Events in Video," to appear in the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04), Washington, DC.

## **6 Contributions**

### **6.1 Contributions within discipline**

A robust technique for tracking multiple people across a camera network

Techniques for auto-elimination of specified human subjects from video content.

Semi-automated system for rapid (10% of real-time) viewing, annotating and truthing of continuously captured video.

### **6.2 Contributions to other disciplines**

Initial field studies were exploratory, but the subsequent planned CareMedia/nursing home collaboration will result in the world's most comprehensive study of the ecology of behaviors in a dementia ward to date.

### **6.3 Contributions to human resource development**

Nursing home field studies have educated computer scientists with respect to issues of privacy and sensitivity in automated analysis of observational data.

### **6.4 Contributions to resources for research and education**

Nothing significant yet.

## **6.5 Contributions beyond science and engineering**

Nothing significant yet.

## **7 Special Requirements**

### **7.1 Objectives and Scope**

A brief summary of the work to be performed during the next year of support if changed from the original proposal.

No change.

### **7.2 Special Reporting Requirements**

Do special terms and conditions of your award require you to report any specific information that you have not yet reported?

No.

### **7.3 Unobligated Funds**

Do you anticipate that more than twenty percent of the funds under your NSF award will remain unobligated at the end of the period for which NSF currently is providing support?

No.

### **7.4 Animals, Human Subjects, Biohazards**

Has there been any significant change in animal care and use, use of human subjects, or biohazards, from what has previously been approved?

No.