

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

**NSF Cooperative Agreement No. IIS-0205219
Annual Progress Report
April 2006**

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

Principal Investigator: Howard Wactlar
School of Computer Science

Co-PI's:

Alexander Hauptmann Computer Science Dept.	Takeo Kanade Robotics Institute
Scott Stevens Computer Science Dept.	Ashok Bharucha 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
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

Asbury Heights 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 work through the privacy, consent, and data protection concerns associated with placing surveillance audio and video equipment in the nursing facilities on a home-by-home basis. We have received Carnegie Mellon University Institutional Review Board approval to conduct such research investigations with human subjects, as well as approval from the State Board of Health. The experiments at the nursing facilities themselves are conducted under the supervision of Dr. Ashok Bharucha, M.D., a geriatric psychiatrist with the CareMedia project. Our Advisory Board for the project gauges ethical concerns and assesses broader societal implications for patient data collection and use in controlled settings.

We have continued to experiment on preliminary “surveillance-type” video, gathered during a pilot study at a nursing care facility, Longwood at Oakmont’s Dementia Care Facility. In order to automatically label social interaction events, we employ some physically measurable parameters extracted from video and audio signals, such as location, motion, shape, color, and audio signal power, to describe coarse and fine motor actions as well as activities. These parameters address physical coding in our scheme and are used together with twenty social events to construct the hierarchical structure of social interaction events. We use a dynamic Bayes network, SVM, Gaussian mixture models and HMM to automatically detect and label the social interaction events under the interval-coding scheme.

In our preliminary studies, we found that the physical measures of location of people, hand motion, talking, and temporal consistency information are ranked as high priorities in the task of detecting interactions. At the social level, three individual events: "talking", "walking", and "leaving", are the most important ones to predict and label social interactions. Gross motor related events, such as "walking", "approaching", and "leaving" can be detected with a precision of about 90%, while fine motor related events such as hand-in-hand and the audio channel event "talking" can be detected with a precision of less than 60%. Many complex social interaction events, e.g., "standing and talking", "walking together", "passing", can be referenced from low-level events with precisions from 86% to 96%.

In analyzing the pilot study recordings, eight consenting residents were observed in the non-private spaces of the dementia unit for 13.6-24.6% of the recorded time. Of this, interpersonal interactions of any kind (verbal or physical; with staff, visitors, other residents) were noted in less than 20% of the recorded time. Meal times accounted for more than 75% of the interpersonal interactions. In comparison to standardized on-site observer rating results, the automatic system captured and tracked seven serious aggressive interactions in the period of observation, preceded in four cases by verbal aggression within 30 seconds of the events, while human observers only reported three of them with the conventional rating scale. Six elopements from the unit behind unsuspecting staff and visitors were also unnoticed by staff.

In late 2005, we recorded a 25-day continuous period in a single ward of the Longwood nursing home in a second CareMedia clinical deployment. Based on recordings and analysis of the pilot study at the same facility in 2002, we instrumented the space to record 24 hours a day with 23 cameras for 25 days. The recording took place in the public areas of the home (see Figure 1), including hallways, entrance areas, dining area, and the recreation room. The 23 640x480 pixel resolution cameras recorded directly into MPEG-2 format, in individual files of 15 minutes duration, to reduce the effects of system glitches. In total, we recorded about 13,000 hours of video (35 terabytes of MPEG-2 files). Together with the video data, we obtained permission to access the medical records of 14 seniors in residence at the facility at that period. Analysis of the data will continue over the coming months

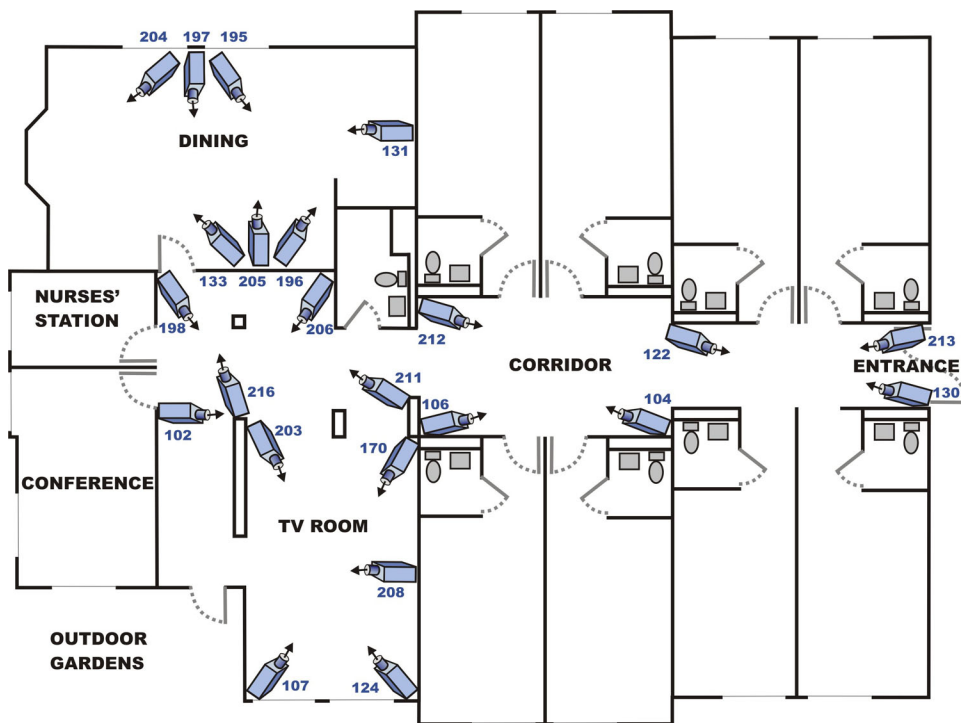


Figure 1: Floor plan of second CareMedia deployment showing camera placement.

2.1.1 Automated Video and Sensor Analysis for Geriatric Care

The Omnibus Budget Reconciliation Act of 1987 (OBRA-87) mandated non-pharmacological interventions for the emotional and behavioral problems of nursing home residents prior to resorting to pharmacotherapy. Nonetheless, the design and implementation of specific non-pharmacological interventions has been hampered by the lack of a sound understanding of the environmental context of these disturbances in nursing homes. CareMedia is developing a suite of techniques and methodologies that capture in real-time a continuous audiovisual and sensor record of the pattern of activities, behaviors and social interactions of nursing home residents while simultaneously developing tools for automated data reduction, and safeguarding privacy.

Application of computer machine intelligence technologies in a nursing home environment is demonstrated as feasible, and will permit *truly longitudinal*, ecological studies of chronic care populations with continuous capture of *objective* measurements of their patterns of activities, behaviors and social interactions. Objective, continuous measurement of various aspects of physical and emotional health of nursing home residents will permit more appropriately targeted interventions, and this technology will serve as a sensitive outcomes measurement tool.

2.1.2 Detecting Social Interaction in a Nursing Home Environment

Social interaction plays an important role in our daily lives. It is one of the most important indicators of physical or mental impairment of aging patients. We conducted a “Wizard of Oz” study on the feasibility of detecting social interaction with sensors in skilled nursing facilities. Our study explored statistical models that could be constructed to monitor and analyze social interactions among aging patients and nurses. We were also interested in identifying sensors that might be most useful in interaction detection; and determining how robustly the detection can be performed with noisy sensors.



Figure 2. Examples of interaction patterns in a nursing home.

We simulated a wide range of plausible sensors using human labeling of audio and visual data. Based on these simulated sensors, we built statistical models for both individual sensors and combinations of multiple sensors using various machine learning methods. Comparison experiments were conducted to

demonstrate the effectiveness and robustness of the sensors and statistical models for detecting interactions.

2.1.3 Label Disambiguation and Sequence Modeling for Identifying Human Activities from Wearable Physiological Sensors

With recent advances in sensor technology, wearable physiological sensors become commonplace and play a vital role in health care. Wearable physiological sensors can provide a faithful record of a patient's physiological states without constant attention of caregivers. A computer program that can infer human activities from physiological recordings will be a valuable tool for physicians. In this paper we investigate to what extent current machine learning algorithms can correctly identify human activities from physiological sensors. We further identify two problems that developers need to address. The first problem is the labels of training data are inevitably noisy due to difficulties of annotating thousands of hours of data. Segments labeled as watching TV may contain other activities, and the wearer may forget to annotate while watching TV. The second problem lies in the continuous nature of human activities, which violates the independence assumption between samples made by many classifier learning algorithms. We approach the first problem of noisy labeling in the multiple-label framework. To address temporal dependence we develop SVM-based Markov Models for continuous recordings that take context into consideration. We evaluated the methods on 12,000 hours of the physiological data recordings from BodyMedia armbands.

2.1.4 Online Learning of Region Confidences for Object Tracking

We developed an online learning method for object tracking. Motivated by the attention shifting among local regions of a human vision system during tracking, we allowed different regions of an object to have different confidences. The confidence of each region is learned online to reflect the discriminative power of the region in feature space and the probability of occlusion. The distribution of region confidences is employed to guide a tracking algorithm to find correspondences in adjacent frames of video images. Only high confidence regions are tracked instead of the entire object. We demonstrated the feasibility of this method in video surveillance applications. The method can be combined with many other existing tracking systems to enhance robustness of these systems.

2.1.5 Modeling Background from Compressed Video

Background models have been widely used for video surveillance and other applications. Methods for constructing background models and associated application algorithms are mainly studied in spatial domain (pixel level). Many video sources, however, are in a compressed format before processing. In this effort, we developed an approach to construct background models directly from compressed video data. Our approach utilizes the information exploited from DCT coefficients at block level to construct accurate background models at pixel level. We implemented three representative algorithms of background models in

the compressed domain, and theoretically explored their properties and the relationship with their counterparts in the spatial domain. We subsequently introduced some general technical improvements to make the approach more capable for a wide range of applications. The method can achieve the same accuracy as the methods that construct background models from the spatial domain with much lower computational cost (26% on average) and more compact storage.

2.1.6 Detecting Head/Shoulders for Person Classification

People detection without certain context, such as a known background, is generally a difficult problem. We designed a strategy that evaluates all possible regions in an image. In this case, the computation cost becomes an important issue, especially if want to apply more informative features, and use complex classifiers. For example, using a well-trained SVM with hundreds of complex image features can cost hours to process just one frame of image with medium size.

With these concerns in mind, the initial system we designed has a cascaded structure. The front end of the system makes use a generative model with simple features for quick filtering, while a subsequent SVM classifier operates on a more complex feature set. The processing time of the system is within seconds per frame, reducing the candidate regions by around 100:1, while maintaining a fair recall rate. Based on the current result, further improvement of precision by using more complex feature /machine learning strategies becomes feasible.

2.1.7 Evaluation and User Studies with Respect to Video Summarization and Browsing

We have investigated “surrogates” that represent a video document or set of video documents in an abbreviated manner. We evaluated the evolution of surrogates such as thumbnails, storyboards, video skims, and video collages. The merits of discount usability techniques for iterative improvement and evaluation were considered, as well as the structure of formal empirical investigations that address the human computer interaction metrics of efficiency (can I finish the task in reasonable time?), effectiveness (can I produce a quality solution?), and satisfaction (would I be willing or eager to repeat the experience again?). The three metrics may be correlated, e.g., an interface that is very satisfying may motivate its user to greater performance and hence higher effectiveness, while conversely an unsatisfying interface may produce extremely slow activity leading to poor efficiency. These three usability aspects were judged in greater detail as they relate to HCI research in general, with the conclusion that all three are necessary to get an accurate assessment of an interface’s usability.

2.1.8 Visual Vocabulary Construction for Mining Biomedical Images

We explored how to succinctly describe the characteristics of multiple images, given a large collection of medical images of several conditions and treatments. How can data mining help biologists focus on important regions in the images or on the differences between different experimental conditions? If the images were text documents, we could find the main terms and concepts for each condition by existing IR methods (e.g., tf/idf and LSI). We explored something analogous, but for the much more challenging case of an image collection: We developed a visual vocabulary automatically by breaking images into $n \times n$ tiles and deriving key tiles ("ViVos") for each image and condition.

We experimented with numerous domain-independent ways of extracting features from tiles (color histograms, textures, etc.), and several ways of choosing characteristic tiles (PCA, ICA). We performed experiments on two disparate biomedical datasets. The quantitative measure of success is classification accuracy: Our "ViVos" achieved high classification accuracy (up to 83% for a nine-class problem on feline retinal images).

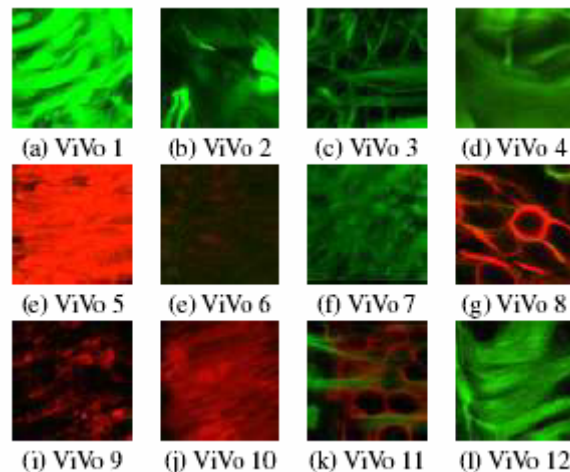


Figure 3. The vocabulary is automatically constructed from a set of images.

More importantly, qualitatively, our "ViVos" did an excellent job as "visual vocabulary terms": they have biological meaning, as corroborated by domain experts; they help spot characteristic regions of images, exactly like text vocabulary terms do for documents; and they highlight the differences between pairs of images.

2.2 Findings

Bharucha, A. J., Pollock, B.G., Dew, M.A., Atkeson, C., Stevens, S., Wactlar, H., Chen, D., "CareMedia: Automated Video and Sensor Analysis for Geriatric Care," Proceedings of the Fifth Annual WPIC Research Day, University of Pittsburgh School of Medicine, June 13, 2005. Also presented under the panel, Innovations in Geriatric Care (Co-Chairs: Mark Miller, MD, Patricia Arean,

PhD), Annual Meeting of the American Association for Geriatric Psychiatry, San Juan, Puerto Rico, March 10-13, 2006; and as a poster presenting the feasibility pilot data at the Annual Meeting of the American Medical Directors Association, Dallas, TX, March 17, 2006.

CareMedia is developing a suite of techniques and methodologies that capture in real-time a continuous audiovisual and sensor record of the pattern of activities, behaviors and social interactions of nursing home residents while simultaneously developing tools for automated data reduction, and safeguarding privacy. Application of computer machine intelligence technologies in a nursing home environment is feasible, and will permit truly longitudinal, ecological studies of chronic care populations with continuous capture of objective measurements of their patterns of activities, behaviors and social interactions.

Bhattacharya, A., Singh, A.K., Faloutsos, C., Yang, H., Verardo, M.R., Pan, J-Y., Ljosa, V., "ViVo: Visual Vocabulary Construction for Mining Biomedical Images," Fifth IEEE Conference on Data Mining (ICDM'05), Houston, TX, November 27-30, 2005.

Given a large collection of medical images of several conditions and treatments, how can we succinctly describe the characteristics of each setting? For example, given a large collection of retinal images from several different experimental conditions (normal, detached, reattached, etc.), how can data mining help biologists focus on important regions in the images or on the differences between different experimental conditions? If the images were text documents, we could find the main terms and concepts for each condition by existing IR methods (e.g., tf/idf and LSI). We propose something analogous, but for the much more challenging case of an image collection: We propose to automatically develop a visual vocabulary by breaking images into $n \times n$ tiles and deriving key tiles ("ViVos") for each image and condition. We experiment with numerous domain-independent ways of extracting features from tiles (color histograms, textures, etc.), and several ways of choosing characteristic tiles (PCA, ICA). We perform experiments on two disparate biomedical datasets. The quantitative measure of success is classification accuracy: Our "ViVos" achieve high classification accuracy (up to 83% for a nine-class problem on feline retinal images). More importantly, qualitatively, our "ViVos" do an excellent job as "visual vocabulary terms": they have biological meaning, as corroborated by domain experts; they help spot characteristic regions of images, exactly like text vocabulary terms do for documents; and they highlight the differences between pairs of images.

Chen, D., Gao, W., Yang, J., "Online Learning of Region Confidences for Object Tracking," The Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, in conjunction with the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, October 15-16, 2005.

This paper presents an online learning method for object tracking. Motivated by the attention shifting among local regions of a human vision system during tracking, we propose to allow different regions of an object to have different confidences. The confidence of each region is learned online to reflect the discriminative power of the region in feature space and the probability of occlusion. The distribution of region confidences is employed to guide a tracking algorithm to find correspondences in adjacent frames of video images. Only high confidence regions are tracked instead of the entire object. We demonstrate feasibility of the proposed method in video surveillance applications. The method can be combined with many other existing tracking systems to enhance robustness of these systems.

Chen, D., Wactlar, H., Yang, J., "A Study of Detecting Social Interaction in a Nursing Home Environment," International Conference on Computer Vision workshop on Human-Computer Interaction, In conjunction with the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, October 15-16, 2005.

Social interaction plays an important role in our daily lives. It is one of the most important indicators of physical or mental diseases of aging patients. In this paper, we present a Wizard of Oz study on the feasibility of detecting social interaction with sensors in skilled nursing facilities. Our study explores statistical models that can be constructed to monitor and analyze social interactions among aging patients and nurses. We are also interested in identifying sensors that might be most useful in interaction detection; and determining how robustly the detection can be performed with noisy sensors. We simulate a wide range of plausible sensors using human labeling of audio and visual data. Based on these simulated sensors, we build statistical models for both individual sensors and combinations of multiple sensors using various machine learning methods. Comparison experiments are conducted to demonstrate the effectiveness and robustness of the sensors and statistical models for detecting interactions.

Chen, D. and Yang, J., “Robust Object Tracking Via Online Spatial Bias Appearance Model Learning,” Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI).

This paper presents a robust object tracking method via spatial bias appearance model learned dynamically in video. Motivated by the attention shifting among local regions of a human vision system during object tracking, we propose to allow different regions of an object to have different confidences and guide object tracking with a dynamic spatial bias appearance model estimated from region confidences. The confidence of a region is defined to reflect the discriminative power of the region in feature space and the probability of occlusion. We propose a novel hierarchical Monte Carlo (HMC) algorithm to learn the region confidences dynamically in every frame. The algorithm consists of two levels of Monte Carlo processes, implemented using two particle filtering procedures in each level, and can efficiently learn outstanding regions through video frames by exploiting the temporal consistency of region confidences. Various spatial bias models are presented to model the distribution of region confidences and the resulted spatial bias map is employed to adapt the appearance model of the object and to guide a tracking algorithm in searching correspondences in adjacent frames of video images. We demonstrate feasibility of the proposed method in video surveillance applications. The proposed method can be combined with many other existing tracking systems to enhance robustness of these systems

Chen, M-Y., Hauptmann, A., Christel, M., Wactlar, H., “Putting Active Learning into Multimedia Applications: Dynamic Definition and Refinement of Concept Classifiers,” ACM Multimedia 2005, National University of Singapore, Singapore, November 6-11, 2005.

We developed an extensible system for video exploitation that puts the user in control to better accommodate novel situations and source material. Visually dense displays of thumbnail imagery in storyboard views are used for shot-based video exploration and retrieval. The user can identify a need for a class of audiovisual detection, adeptly and fluently supply training material for that class, and iteratively evaluate and improve the resulting automatic classification produced via multiple modality active learning and SVM. By iteratively reviewing the output of the classifier and updating the positive and negative training samples with less effort than typical for relevance feedback systems, the user can play an active role in directing the classification process while still needing to truth only a very small percentage of the multimedia data set. Examples are given illustrating the iterative creation of a classifier

for a concept of interest to be included in subsequent investigations, and for a concept typically deemed irrelevant to be weeded out in follow-up queries. Filtering and browsing tools making use of existing and iteratively added concepts put the user further in control of the multimedia browsing and retrieval process.

Christel, M., "Evaluation and User Studies with Respect to Video Summarization and Browsing," in *Multimedia Content Analysis, Management and Retrieval 2006*, part of the IS&T/SPIE Symposium on Electronic Imaging 2006, San Jose, CA, January 17-19, 2006.

The Informedia group at Carnegie Mellon University has since 1994 been developing and evaluating surrogates, summary interfaces, and visualizations for accessing digital video collections containing thousands of documents, millions of shots, and terabytes of data. This paper surveys the Informedia user studies that have taken place through the years, reporting on how these studies can provide a user pull complementing the technology push as automated video processing advances. The merits of discount usability techniques for iterative improvement and evaluation are presented, as well as the structure of formal empirical investigations with end users that have ecological validity while addressing the human computer interaction metrics of efficiency, effectiveness, and satisfaction. The difficulties in evaluating video summarization and browsing interfaces are discussed. Lessons learned from Informedia user studies are reported with respect to video summarization and browsing, ranging from the simplest portrayal of a single thumbnail to represent video stories, to collections of thumbnails in storyboards, to playable video skims, to video collages with multiple synchronized information perspectives.

Hao, W.-H., and Hauptmann, A., "Label Disambiguation and Sequence Modeling for Identifying Human Activities from Wearable Physiological Sensors," to appear in ICME'06 Special Session: Multimedia Technologies for Healthcare.

In this paper we investigate to what extent current machine learning algorithms can correctly identify human activities from physiological sensors. We further identify two problems that developers need to address. The first problem is the labels of training data are inevitably noisy due to difficulties of annotating thousands of hours of data. Segments labeled as watching TV may contain other activities, and the wearer may forget to annotate while watching TV. The second problem lies in the continuous nature of human activities, which violates the independence assumption between samples made by many classifier learning

algorithms. We approach the first problem of noisy labeling in the multiple-label framework. To address temporal dependence we develop SVM-based Markov Models for continuous recordings that take context into consideration. We evaluate the proposed methods on 12,000 hours of the physiological recordings recorded from BodyMedia armbands and report learned lessons.

Wang, W., Gao, W., Yang, J., Chen, D., “Modeling Background from Compressed Video,” The Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, in conjunction with the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, October 15-16, 2005.

Background models have been widely used for video surveillance and other applications. Methods for constructing background models and associated application algorithms are mainly studied in spatial domain (pixel level). Many video sources, however, are in a compressed format before processing. In this paper, we propose an approach to construct background models directly from compressed video data. The proposed approach utilizes the information exploited from DCT coefficients at block level to construct accurate background models at pixel level. We implemented three representative algorithms of background models in the compressed domain, and theoretically explored their properties and the relationship with their counterparts in the spatial domain. We also present some general technical improvements to make the proposed approach more capable for a wide range of applications. The proposed method can achieve the same accuracy as the methods that construct background models from the spatial domain with much lower computational cost (26% on average) and more compact storages.

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

Bharucha, A.J., “Continuous Audiovisual Capture in a Dementia Unit: Recruitment and Ethical Challenges.” 58th Annual Scientific Meeting of the Gerontological Society of America, Orlando, FL, November 20, 2005.

Bharucha, A. J., Pollock, B.G., Dew, M.A., Atkeson, C., Stevens, S., Wactlar, H., Chen, D., “CareMedia: Automated Video and Sensor Analysis for Geriatric Care,” Proceedings of the Fifth Annual WPIC Research Day, University of Pittsburgh School of Medicine, June 13, 2005. Also presented under the panel, Innovations in Geriatric Care (Co-Chairs: Mark Miller, MD, Patricia Arean, PhD), Annual Meeting of the American Association for Geriatric Psychiatry, San Juan, Puerto Rico, March 10-13, 2006; as a poster at the Annual Meeting of the American Association for Geriatric Psychiatry, San Juan, Puerto Rico, March 12, 2006; and as a poster presenting the feasibility pilot data at the Annual Meeting of the American Medical Directors Association, Dallas, TX, March 17, 2006.

Bhattacharya, A., Singh, A.K., Faloutsos, C., Yang, H., Verardo, M.R., Pan, J-Y., Ljosa, V., “ViVo: Visual Vocabulary Construction for Mining Biomedical Images,” Fifth IEEE Conference on Data Mining (ICDM'05), Houston, TX, November 27-30, 2005.

Chen, D., Gao, W., Yang, J., “Online Learning of Region Confidences for Object Tracking,” The Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, in conjunction with the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, October 15-16, 2005.

Chen, D., Wactlar, H., Yang, J., “A Study of Detecting Social Interaction in a Nursing Home Environment,” International Conference on Computer Vision workshop on Human-Computer Interaction, In conjunction with the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, October 15-16, 2005.

Chen, D. and Yang, J., “Robust Object Tracking Via Online Spatial Bias Appearance Model Learning,” Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI).

Chen, M-Y., Hauptmann, A., Christel, M., Wactlar, H., “Putting Active Learning into Multimedia Applications: Dynamic Definition and Refinement of Concept Classifiers,” ACM Multimedia 2005, National University of Singapore, Singapore, November 6-11, 2005.

Christel, M., “Evaluation and User Studies with Respect to Video Summarization and Browsing,” in Multimedia Content Analysis, Management and Retrieval 2006, part of the IS&T/SPIE Symposium on Electronic Imaging 2006, San Jose, CA, January 17-19, 2006.

Hao, W.-H., and Hauptmann, A., "Label Disambiguation and Sequence Modeling for Identifying Human Activities from Wearable Physiological Sensors," to appear in ICME'06 Special Session: Multimedia Technologies for Healthcare.

Wang, W., Gao, W., Yang, J., Chen, D., "Modeling Background from Compressed Video," The Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, in conjunction with the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, October 15-16, 2005.

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.

Technique for masking the ID of specific individuals in video.

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 and 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

Video capture enabled manually assisted observation of unreported aggressive acts by patients, and unnoticed patient elopements.

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?

Protocols have been revised to permit studies at additional sites and to comply with local requirements. Further approval has been obtained from the Commonwealth of Pennsylvania Department of Health.