

CONTRACTOR'S PROGRESS STATUS AND MANAGEMENT REPORT

Intelligent Collaboration and Visualization

for the period 1 October to 31 December 1998

Report #7
CDRL A001

Contract N66001-97-C-8517

13 April 1999

SUBMITTED TO

Receiving Officer
e-mail address: "icvnrad@nosc.mil"

Rich Laverty
619-553-2918
laverty@nosc.mil

Steve Murray
619-553-6350
murrays@spawar.navy.mil

SUBMITTED BY

Carnegie Mellon University
Pittsburgh, PA 15213

Principal Investigator
Howard Wactlar
(412) 268-2571
fax: (412) 268-5576
hdw@cmu.edu

Administrative Contact
Dale James
(412) 268-7689
dale+@cs.cmu.edu

Contract/Financial Contact
Karen M. Faber
(412) 268-5838
(412) 268-5841
faber+@andrew.cmu.edu

Do not distribute to DTIC or other data depositories.

Distribution authorized to DOD components only; premature dissemination (date). Other requests shall be referred to Naval Command, Control and Ocean Surveillance Center (NCCOSC), RDT&E Division, San Diego, California 92152-5000.

Quarterly Status Report

Intelligent Collaboration and Visualization
for the period 1 October to 31 December 1998
Cooperative Agreement N66001-97-C-8517
CDRL A001

1.0 Purpose of Report

This status report is the quarterly contract deliverable (CDRL A001) which summarizes the effort expended by the Carnegie Mellon University team in support of Intelligent Collaboration and Visualization (IC&V) on Contract N66001-97-C-8517.

2.0 Project Members

Wactlar	20% of effort
Christel	20
Faloutsos	33
Gong	100
Hauptmann	15
Reddy	5
Stevens	35
Other staff:	750%

3.0 Project Description (last modified 2/97)

This work will develop tools, techniques, and systems allowing people to capture a complete record of their personal experiences, and to share them in collaborative settings. Users may range from rescue workers carrying personalized information systems in operational situations to remote crisis managers in coordinating roles. Personal Informedia Experience-on-Demand (EoD) units record audio, video, GPS and other sensory data, which can be annotated by human participants. The EoD environment synthesizes data from many EoD units into a “collective experience” – a global perspective of ongoing and archived personal experiences. Distributed collaborators are thereby brought together over time and space to share meaning and perspectives.

Each constituent EoD unit captures and manages information from its unique point of view, making this information available to others in the EoD environment. Each operates as a portable, interoperable, Informedia system, allowing search and retrieval by both its human operator and remote systems. The EoD environment thus enables integration of the multiple points of view to provide more details for local decision-making and superior event coverage in support of real-time collaboration. The EoD capability will significantly improve situation awareness and analysis, both in real-time and retrospectively. The indexed and summarized information also enables “remembering” analogous past experiences and “projecting” into future simulated ones. Techniques will be developed to manage the vast quantities of information and to search, summarize, and visualize video,

audio, and text content and annotations from multiple perspectives. The foundation for this work, the Informedia Digital Video Library (DVL) Project, has demonstrated the successful application of speech, image, and natural language processing in automatically creating a rich, indexed, searchable multimedia information resource. We will build on these technologies, moving beyond a DVL into new information spaces by addressing collaboration and summarization of multiple simultaneous information generators integrated across people, time, and space.

4.0 Performance Against Plan

Spending proceeded as planned during this reporting period.

5.0 Major Accomplishments to Date

- Integrated VOOCR (video optical character recognition) into EoD's underlying Informedia system.
- Improved our image-retrieval method by exploiting color-cluster and image-region characteristics.
- Tested image-retrieval performance using a larger image set (10,000 images) from the Informedia database.
- Modularized our image-retrieval method to facilitate integrating it into the EoD system.
- Integrated a preliminary data-visualization technique, VIBE, into the EoD system.
- Developed MindReader, a method that improves query by image content.

6.0 Artifacts Developed During the Past Quarter

We developed a technique for motion detection in video using reference frames. This algorithm, while not yet fast enough to include in the EoD system, demonstrates the feasibility of filtering redundant video from extended EoD-captured content.

7.0 Issues

7.1 Open issues with no plan, as yet, for resolution:

- None.

7.2 Open issues with plan for resolution:

- Clearer recording of audio and video with the EoD vest.
- Client does not allow two concurrent users at the time. Must build capability of visualization of two users.

7.3 Issues resolved:

- Rudimentary filtering of redundant video, and composition of basic panoramic views.

8.0 Near-term Plan

- Incorporate into the system the ability to acquire and process digital-compass data, in addition to GPS. This new data type will allow tracking view direction.
- Refine the color-space representation used in our color-clustering technique to better support real application imagery.
- Move the system to an Oracle database for greater speed and reliability.
- Conduct more experiments with two or more concurrent EoD users in more realistic situations.

9.0 Completed Travel

<>

10.0 Equipment Purchases and Description

<>

11.0 Summary of Activity

Work Focus:

- Generalized our VOOCR technique to handle both light text on dark background and the reversed case. Processing text regions appropriately, according to their relative foreground/background luminance, allows the system to "read" overlaid text in wider range of colors.
- Reimplemented our heuristic, color-clustering technique on a larger scale — 4× higher resolution and more numerous images, 10K vs. 80. While considering region shape and location in addition to color, this method of determining image similarity by content is already significantly faster than the more conventional histogram approach.
- Completed work on the named-face technique that can automatically label people appearing in broadcast news video. The system builds a reliable database of name/face correspondences, verifying and refining the resource by analyzing external Web-site information.
- Developed a fully automatic technique that can extract unlimited training data from closed-captioned television broadcasts. Testing with 111 hours of data thus obtained improved our best Sphinx III acoustic model by 7.6% on DARPA's Hub4 broadcast-news evaluation task.
- Implemented a selective training strategy that employs such unlimited data efficiently by extracting a critical subset and training on underrepresented triphones. Using only 60% of the 111 hours data, we reduced training costs and improved the relative word-error rate 6.8% over the baseline Hub4 system, almost as well as with the larger training corpus and better than a similar amount of unselected, manually transcribed data.
- Developed a technique that exploits closed captions to enable recognizing nonspeech sounds. By automatically locating 110 instances of the [LAUGHTER] noise-token in the

Informedia database, manually trimming their boundaries, and training with a new, corresponding phoneme, we improved the recognizer's word-error rate by 2.7%, relative to the baseline.

- Demonstrated the feasibility of determining current position within a building, where GPS data are not available, by interpolating access-point locations within the wireless-network infrastructure.
- Removed inconsistent geocode data within the DB, improving accuracy and enabling a graphic version of cursor text-following: automatic display of maps covering locations mentioned in audio track.
- Improved information visualization by creating interface sliders that allow the user to adjust filtering criteria for date, segment size, and query relevance score.
- Improved the Informedia display, based on analytic and pilot user studies, to better use screen "real estate" and to facilitate user control of dynamically adjustable video abstractions:
 - Exploit context established by the user's query, recognized speech, and speech/image alignment to select candidate filmstrip images
 - Assemble skim components to emphasize query-matches and display at a user-controlled compression ratio.
- Developed a new version of the Informedia search engine that will provide greater speed and flexibility by using database-resident indexes instead of separate files.
- Validated the Query Manager's scalability. Testing with ten users on "thin client" machines demonstrated effective query interactions with back-end servers.
- Developed a QC tool that will help ensure DB integrity.

11.1 Significant Events:

- None this period.