CONTRACTOR'S PROGRESS STATUS AND MANAGEMENT REPORT

Intelligent Collaboration and Visualization

for the period 1 October to 31 December 1999
Report #11
CDRL A001

Contract N66001-97-C-8517

31 December 1999

SUBMITTED TO

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

Rich Laverty	Steve Murray
619-553-2918	619-553-6350
laverty@nosc.mil	murrays@spawar.navy.mil

SUBMITTED BY

Carnegie Mellon University
Pittsburgh, PA 15213

Principal Investigator	Administrative Contact	Contract/Financial Contact
Howard Wactlar	Dale James	Karen M. Faber
(412) 268-2571	(412) 268-7689	(412) 268-5838
fax: (412) 268-5576		(412) 268-5841
hdw@cmu.edu	dale+@cs.cmu.edu	faber+@andrew.cmu.edu

Quarterly Status Report

Intelligent Collaboration and Visualization
for the period 1 October to 31 December 1999
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
Christel
Faloutsos
Hauptmann
Reddy
Stevens
Other technical staff

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 VOCR (video optical character recognition) into EoD's underlying Informedia system, allowing EoD to "read" some naturally occurring scene text.
- Improved our image-retrieval method by exploiting color-cluster and image-region characteristics.
- Tested image-retrieval performance using a larger image set (20,000 images) from the Informedia database.
- Modularized our image-retrieval method to facilitate integrating it into the EoD system.
- Developed an interactive map-based display and query mechanism enabling better access and enhanced visualization of continuously captured and archival library data.
- Developed a method to synthesize a basic panoramic view from multiple perspectives that might be provided by concurrent EoD users.
- Developed a technique for motion detection in video, by comparing against reference frames.
- Automatic silence filtering integrated into video processing.
- Location names can now be disambiguated automatically in geocoding.

6.0 Artifacts Developed During the Past Quarter

- None this quarter.

7.0 Issues

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

- The granularity of the digital compass data proved too coarse to allow precise readings for panoramic reconstruction, thus we are forced to reconsider its utility.

7.2 Open issues with plan for resolution:

- More useable, real-world detection and reading of incidental "scene text." Our VOCR can detect some scene text (signs on a truck, for instance) if it has certain properties (horizontal, stable, etc.)
- Enable collection of real-time multiple EoD views from several units.

7.3 Issues resolved:

- Improved scene text detection with backward propagating neural nets.
- Designed a more fault-tolerant EoD architecture.

8.0 Near-term Plan

- Refine and optimize distribution of the IDVLS processing and serving across a number of machines to determine what configuration provides greater utility.
- Develop new strategies for deployment and coordination of multiple, concurrent EoD units.
- Develop new visualization techniques

9.0 Completed Travel

November 1999 – DARPA IC&V Program PIs Meeting, San Diego, CA – Howard Wactlar and Alex Hauptmann

10.0 Equipment Purchases and Description

Data projector - $5,664

11.0 Summary of Activity

We integrated a digital compass into the EoD system to determine in which direction the user is moving and/or facing. Unfortunately, the granularity of the captured data proved too coarse to allow for panoramic reconstruction from multiple independent views. We are forced to reconsider the utility of digital compass data.

In an attempt to improve VOCR, we applied color clustering techniques to a set of video images. Our experiments resulted in substantial performance loss compared to our original OCR processing. Our best results came from using an extra three pixel boundary in text images fed into the original OCR package. This yielded a 40% character error rate per image, compared to 51% previously. Contrary to expectations, widening the boundary to five pixels increased error rate to 45%.

We are now training neural nets for scene text detection. The backward propagation neural network is trained with 16x16 pixel tiles (taken from grayscale images) which are either text or non-text tiles. The algorithm consists of taking the FFT (fast fourier transform) of each tile and feeding this to the neural net. The neural net is then trained to learn the characteristics of tiles containing text and those that do not contain text. We discovered that we can detect (but not recognize) text of different styles and orientations, eliminating text appearance constraints imposed by our previous, non-neural net text detection processing. Preliminary results show approximately 80% accuracy for scene text detection on the training data.

We have continued investigating multimodal queries, specifically, searching with text while also searching indicated geographic areas. A weighting scheme has been developed to combine results produced from these different modalities, with direct manipulation interfaces provided so that the user can adjust the weights to match specific needs. In addition, features have been extracted from the result set for use in relevance feedback, whereby the user informs the system which results are relevant to his needs. The features common to those results marked as relevant are used to reorder the result set accordingly. Experiments are being planned to measure the effectiveness of multimodal map and text queries and relevance feedback strategies, as compared to video information retrieval interfaces with text-only query, and text query with relevance feedback.

We set up an ad-hoc wireless LAN, using Lucent's new IEEE 802.11 wireless cards and mid-range antennae, and connected this LAN to the campus network both through the wireless-Andrew network and through the normal wired network. This was achieved by using a software router with Network Address Translation (NAT) capabilities. This LAN operates over the upgraded Wireless Andrew at 11 Mbps, compared to the shared 2Mbps that Wireless Andrew provided us last quarter. Moreover, by using the antennae as mentioned above, we were able to have significantly larger coverage outdoors (~300 feet, while keeping close to peak bandwidth).

We designed an EoD architecture that would be more fault-tolerant, enabling recovery after lost connectivity. The system would store collected data locally, and continue forwarding this data to our outside server when the connection is restored. To this end, we have implemented a “data-relocator” which would receive a signal from the “data-collector” when new data arrives, and then move this data over to the remote server in the outside world (beyond the limits of the wireless LAN). Moreover, the data-relocator would constantly monitor the network to determine whether it was able to transport the data. In case of failure during transmission (namely, a physical network disconnection), the data-relocator would attempt full retransmission of the interrupted data when the network again becomes available.

11.2 Significant Events:

- Attended, presented and demonstrated current system at the DARPA IC&V Program PIs Meeting in San Diego, California in November 1999.