

**Informedia-II:
Auto-Summarization and Visualization
Over Multiple Video Documents and Libraries**

**NSF Cooperative Agreement No. IIS-9817496
Semiannual Progress Report
September 2000**

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

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

Co-PI's: **Michael Christel**
Computer Science Dept.

Takeo Kanade
Robotics Institute

Christos Faloutsos
Computer Science Dept.

John Lafferty
Computer Science Dept.

Alexander Hauptmann
Computer Science Dept.

Yiming Yang
Language Technologies Inst.

1 Overview

The Informedia-II Project will change the paradigm for accessing digital video libraries through meaningful, manipulable overviews of video document sets, multimodal queries, and adaptive summarizations of very large amounts of video from heterogeneous distributed sources.

Video information collages are the key technology in Informedia-II and will be built by advancing information visualization research to effectively deal with multiple video documents. A video information collage is a presentation of text, images, audio, and video derived from multiple video sources in order to summarize, provide context, and communicate aspects of the content for the originating set of sources. The collages to be investigated include chrono-collages emphasizing time, geo-collages emphasizing spatial relationships, and auto-documentaries which preserve video's temporal nature. Users will be able to interact with the video collages to generate multimodal queries across time, space, and sources. Video collages are made adaptive by giving preference to the concepts and query terms in the user's interaction history. The synthesis and summarization functions underlying these collages will be made possible through extensions of text clustering and expectation maximization algorithms to video and audio features.

2 Research and Testbed Summary

2.1 *Video Information Collages: Adaptive Visualization and Summarization*

2.1.1 Contextual term phrases in video documents

Continuing from the work of concept extraction in text started from last reporting period, we investigated the use of contextual term phrases to extend the summarization and presentation dimensions to represent video documents. In addition to title and topics assigned to each video document, a document is represented by a short list of most-important phrases, which provides a quick glance of the diverse content covered in a document.

For more details on the research in contextual information, please refer to [Ng2000]. For more details on the use of contextual term phrases, please refer to [Zhong2000].

2.1.2 Adaptive query-based and adjustable filmstrips

We developed a series of storyboard interfaces with added transcript text features. These interfaces were used in a controlled experiment focusing on the utility of transcript text in storyboards for news video navigation. We wished to explore whether such text resulted in improvements in video navigation, and, if so, whether the amount of text and its synchronization with video imagery affected the navigation task. The text-augmented storyboards performed significantly better than storyboards with no text. Full transcript text produced benefits when presented as a block, whereas reduced contextual text descriptions produced benefits when aligned with storyboard image rows.

These results are interpreted along with subjects' post-experiment rankings in, and have implications for the design of the next generation of digital video players and browsers. [Christel,Warmack2001], submitted for publication, provides more information.

2.1.3 User's interaction history

One goal of Informedia-II is to develop *adaptive* video information collages that provide more focused information of relevance to the user given his or her context. We have now incorporated an interaction history into the IDVLS, with user control over clearing the history, the size of the history cache, and annotations for documenting history items.

The history mechanism has been tested and refined through user studies with CMU students. This component is able to create user sessions, and record all queries from all sessions. Users are able to prioritize queries using features other than time features according to the history.

2.2 Information Analysis and Semantic Summarization

2.2.1 Query expansion with contextual term phrases

The result of co-occurrence analysis on contextual phrases provides the basis for the research work of query expansion in our Informedia client development. Instead of gathering single-word keywords to expand a query, we use specific and contextually relevant term phrases to expand users' query. Our implementation of query expansion provides two modes for operation – automatic and user-specific query expansion. Automatic query expansion will automatically include the ten most relevant term phrases to the original query. User-specific expansion allows users to examine and select a list of relevant term phrases to expand or even change their query.

For more details, please refer to [Zhong2000].

2.2.2 Automatically generating meaningful titles

The problem of title generation involves finding the essence of a document and expressing it in only a few words. The results of a query to the Informedia Digital Video Library are summarized through an automatically generated title for each retrieved news story. Errorful documents such as speech-recognized broadcast news stories present an even greater challenge. We implemented a set of title word selection strategies and evaluated them on an independent test corpus of 579 broadcast news documents, comparing manual transcription results to automatically recognized speech using the CMU Sphinx speech recognition system with a 64000-word broadcast news language model. Using a training collection of 21190 transcribed broadcast news stories, we trained several systems to produce appropriate title words. Overall results show that title generation for speech recognized news documents is possible at a level approaching the accuracy of titles generated for perfect text transcriptions. One surprising phenomenon is that extractive approaches perform slightly better for speech recognized documents than for manual transcripts.

See [Jin,Hauptmann2000] for more information.

2.2.3 Learning spectral image segmentation in video

Good image segmentation of a video is essential for recognizing and indexing objects in video. We are advancing our research in two directions. First, we have deepened our understanding of the grouping and segmentation process in a statistical framework. We have presented a new view of grouping by pairwise similarities. By interpreting the similarities as edge flows in a Markov random walk, and studying the eigenvalues and eigenvectors of the walks' transition matrix, we showed that the equivalence between the steady-state of the Markov random walks and segmentation computed by the Normalized cut criteria. Second, with this statistical interpretation, we are developing algorithms that will allow us to learn global grouping patterns efficiently from local similarity measures. Such algorithms will allow the user to define a set of desired grouping patterns (e.g., rounded convex objects), and have the segmentation system automatically learn to segment those patterns. For details, see [Maila,Shi2000], [Maila,Shi2001].

2.2.4 Context analysis: concept association

We have extended our research in context analysis from associating concepts such as important term phrases and place names to associating concepts extracted from image analysis. In the Informedia Digital Video Library, each news story is represented by a set of media-rich concepts, which can be found in transcript (term phrases, place names, person names, and organization names) and in video stream (still and moving images). We would like to use contextual information to associate text labels to corresponding objects extracted from still and moving images. The result will expand searching capability across multimedia objects in the digital video corpus – visual objects can be retrieved through their automatically associated text labels and, in turn, visual objects can be used to query other information in the library.

We are investigating the suitability of various image similarity functions. An image similarity function provides the basis for identifying similar objects in multiple video documents. We are also implementing the co-occurrence analysis for both image and textual information.

For more details on the analysis and design of context analysis in digital library, please refer to [Ng2000]. For more details on image analysis, please refer to [Maila,Shi2000].

2.3 Multimodal Query: Beyond Query/Browse by Text to Video Exploration

2.3.1 Shape-based object matching in cluttered images

We are developing algorithms for detecting arbitrary objects in cluttered images. In this system, a user will sketch the object (s)he is looking for, and the algorithm would be able to detect such an object in an image at any location, size, and rotation. The underlying shape-based object matching algorithm addresses two key issues. First, we studied how to represent the shape of an object, which will capture both the global geometrical arrangement as well as local appearance. We have developed a self-adaptive shape-context representation, which embodies both the geometrical and statistical nature of object shapes. Second, we have developed a fast algorithm based on divide-and-conquer for finding the best match of objects in this representation. Research efforts are

underway to integrate the image segmentation system into this object matching algorithm. For details, see [Shi, Hampton 2000].

2.3.2 Relevance feedback for a multi-terabyte digital video library

One strategy for addressing the imprecision in search results from large libraries is to use relevance feedback, where the user refines the result set by reissuing more focused queries reflecting the user's particular needs. Through such query reformulation, the intent is for the result set to become increasingly more precise, including only those documents of relevance to the user's information needs.

Two facilities for accomplishing relevance feedback were recently added to the Informedia library interface for text document retrieval; an opaque interface and a penetrable interface. In the opaque interface, the functionality of the relevance feedback component is hidden from the user. In the penetrable interface, the user is shown the underlying information about the functioning of the relevance feedback mechanism and is given the ability to manipulate the output of the relevance feedback component prior to query evaluation. The user has much greater control, but at the expense of more complex interfaces and interactions. We conducted a pilot study with six users of these two interface types for video document retrieval, which suggested areas for further investigation and improvement in the video retrieval domain. See [Zhong2000] for more on relevance feedback.

2.3.3 Synthesized geo-spatial display and query

We enhanced Informedia's map interface to allow *zooming*, with the user choosing level of details ranging from city, country, to region. We added interface functionality to query by administrative entity (e.g., US state or Canadian province), and to show visualizations by administrative entity as one form of such a zoomable map interface with changing levels of detail.

The real value added is location information for video segments that previously had little or none. Every news story does not have an embedded map that becomes part of the broadcast, but through our geocoding, maps can be automatically produced to reflect the areas discussed within each story. Another benefit is that the user can interact with the interface map using the toolbar icons to get additional detail, whereas no such interaction is possible with an image of a map encoded as part of the video stream.

The maps accompanying videos are animated in synchronization with video playback. As places are discussed, they are highlighted on the map. For countries and administrative areas such as states or provinces, the areas contained within their respective polygon boundaries are highlighted by changing the fill color. A glance at the map can then show the areas of current focus. See [Christel,et.al.2000] for more information.

2.3.5 Density biased sampling

Data mining in large data sets often requires a sampling or summarization step to form an in-core representation of the data that can be processed more efficiently. Uniform random sampling is frequently used in practice and also frequently criticized because it will miss small clusters. Many natural phenomena are known to follow Zipf's distribution and the

inability of uniform sampling to find small clusters is of practical concern. Density Biased Sampling probabilistically under-samples dense regions and over-samples light regions. A weighted sample is used to preserve the densities of the original data. Density biased sampling naturally includes uniform sampling as a special case. A memory efficient algorithm is proposed that approximates density biased sampling using only a single scan of the data. We empirically evaluate density biased sampling using synthetic data sets that exhibit varying cluster size distributions finding up to a factor of six improvement over uniform sampling.

More generally, density biased sampling offers a representative sample of the data that includes more of the unexpected points. Any algorithm that does not require that all inputs be distinct can be trivially extended to support a weighted sample. Many statistical algorithms use multiple samples to reduce variability.

Density biased samples should reduce the variability of the algorithms because we can include more of the "unusual" points (i.e., the points that are likely to induce variability) while ensuring a representative sample. It should be possible to efficiently construct a density biased sample using an R-tree index by descending in the R-tree only as far as needed to compute the bin sizes. Using an existing index may make it possible to construct samples without reading the entire database.

For more information, please see [Palmer,Faloutsos2000].

2.3.6 Feedback adaptive loop for content-based retrieval

We have developed a novel relevance feedback query method that is designed to handle disjunctive queries within metric spaces. The user provides weights for positive examples; our system learns the implied concept and returns similar objects. Our method differs from existing relevance-feedback methods that base themselves upon Euclidean or Mahalanobis metrics, as it facilitates learning even disjunctive, concave models within vector spaces, as well as arbitrary metric spaces. In addition, our method is completely example-driven, and imposes no requirements upon the user for other aspects such as feature selection.

Our main contributions from this research are twofold. Not only do we present a novel way to estimate the dissimilarity of an object to a set of desirable objects, but we support it with an algorithm that shows how to exploit metric indexing structures that support range queries to accelerate the search without incurring false dismissals. Our empirical results demonstrate that our method converges rapidly to excellent precision/recall, while outperforming sequential scanning by up to 200%.

Please refer to [Wu,et.al.2000] for further information.

3 Notable Outreach and Inclusion Activities

- The Andy Warhol Museum recently established a collaborative effort with the Informedia project to apply its technology to archiving the complete film and video works (finished productions and raw footage) of Andy Warhol.
- TalkBank, a multimedia database of communicative interactions, was awarded by NSF in the last year with Informedia as an underlying technology enabling the

research of psychologist Brian MacWhinney and University of Pennsylvania's linguist Mark Liberman and computer scientist Peter Buneman. The goal of TalkBank is the creation of a distributed, web-based data archiving system for transcribed video and audio data on communicative interactions. These interactions will include mothers talking with their children, family dinner table talk, classroom interactions, signed language, formal debates, phone calls, talk with foreigners, club meetings, and dozens of other types of communicative interactions. TalkBank will facilitate comparisons across social groups, languages, and situations. The initiative establishes an ongoing interaction between computer scientists, linguists, psychologists, sociologists, political scientists, criminologists, educators, psychiatrists, and anthropologists

- Working with the commercial derivatives offered by MediaSite, an Informedia spin-off endeavor, JTASC (USJFCOM Joint Training, Analysis & Simulation Center) has selected Informedia technology to power its video information and training digital library.
- NASA has selected Dreamtime to make its vast collection of imagery and video available to the masses through electronic access. Working through MediaSite, Dreamtime has indicated very strong interest in working with Informedia technology to index and access this national resource.
- General Motors Research Lab is interested in using a version of Informedia for location based video retrieval in cars. Potential uses range from trip-planning to guided tours or on-demand entertainment. Meetings will continue through September and beyond. These talks are still in the brainstorming stage, but have yielded interesting development ideas.
- Alex Hauptmann taught the 8th ELSNET European Summer School in Language and Speech Communication (TeSTIA2000), 15-30 July 2000. Plenary session on Multimedia Digital Libraries. For more info:<http://www.ilsp.gr/testia/testia2000.html> The course covered issues involving capture, processing, compression, storage, indexing, search, and retrieval of various kinds of audio, video and image media. The intent was to provide a conceptual and technical framework for multimedia digital libraries.

4 Journal and Conference Proceeding Publications

Christel, M.G., Olligschlaeger, A.M, and Huang, C. "Interactive Maps for a Digital Video Library". IEEE MultiMedia 7(1): 60-67, 2000.

Christel, M.G., Warmack, A. "Does Text added to Storyboards Improve Video Navigation?" Submitted for consideration for publication at the ACM CHI 2001 conference in April, 2001.

Faloutsos, C., Traina Jr., C., Traina, A., Seeger, B., "Spatial Join Selectivity Using Power Laws", In Proceedings of SIGMOD 2000, Dallas, TX, May 14-19, 2000.

Jin, R., Hauptmann, A. "Title Generation for Spoken Broadcast News using a Training Corpus", To appear in proceedings of ICSLP 2000, 6th International Conference on Spoken Language Processing, Beijing, China, October 16-20, 2000.

Kennedy, P.E., Hauptmann, A., "Automatic Title Generation using EM", Submitted for publication in ACM Digital Libraries 00, San Antonio, Texas, June, 2000

Maila, M., Shi, J. "Learning Spectral Image Segmentation with Random Walks", Accepted and to appear in Neural Information Processing System(NIPS) 2000, November, 2000.

Maila, M., Shi, J. "A Random walks view of spectral segmentation", Accepted and to appear in 8th International Workshop on Artificial Intelligence and Statistics, January, 2001.

Ng, T.D., "A Concept Space Approach To Semantic Exchange," Ph.D. Dissertation, The University of Arizona. April, 2000.

Palmer, C.R., Faloutsos, C., "Density Biased Sampling: An Improved Method for Data Mining and Clustering", In Proceedings of SIGMOD 2000, Dallas, TX, May 14-19, 2000.

Riedel, E., Faloutsos, C., Ganger, G., Nagle, D. "Data Mining on an OLTP System (Nearly) for Free", ACM SIGMOD, International Conference on Management of Data, Dallas, Texas, May 14-19, 2000.

Shi, J., Hampton, G. "Shape-based Object in a Cluttered Environment", CMU Robotics Internal Report, August, 2000.

Traina Jr., C., Faloutsos, C., Seeger, B., Traina, A., "Slim-trees: High Performance Metric Trees Minimizing Overlap Between Nodes", In Proceedings of the International Conference on Extending Database Technology EDBT 2000, Konstanz, Germany, March 27-31, 2000.

Traina Jr., C., Faloutsos, C., Traina, A., "Distance Exponent: A New Concept for Selectivity Estimation in Metric Trees", Proceedings of the IEEE 16th Intl. Conference on Data Engineering in San Diego, CA, February 29 - March 3, 2000.

Wactlar, H., "Informedia - Search and Summarization in the Video Medium", Proceedings of Imagina 2000 Conference, Monaco, January 31 - February 2, 2000.

Wu, L., Faloutsos, C., Sycara, K., Payne, T. "FALCON: Feedback Adaptive Loop for Content-Based Retrieval", VLDB 2000, Cairo Egypt (to appear), September, 2000.

Yi, B., Faloutsos, C. "Fast Time Sequence Indexing for Arbitrary Lp Norms", VLDB 2000, Cairo, Egypt (to appear), September, 2000.

Zhong, Y., "Apply Multimodal Search and Relevance Feedback In a Digital Video Library", Thesis for the degree MSc. in Information Networking, May, 2000.

5 Presentations, Demonstrations, and Industry Visitors

March 6, 2000: Ryuichi Hanaoka Takaoko, President, Hitachi Cable.

March 9, 2000: System demo and discussion at the Carnegie Museum Pittsburgh I-Net Celebration in the Carnegie Museum of Natural History.

March 14, 2000: Joan Shigekawa, Associate Director, Arts and Humanities, The Rockefeller Foundation. Research discussion/demo.

April 7, 2000: Bob Ross, Vice President for East Coast, CBS. Research discussion/demo.

April 18, 2000: Maureen McFalls, <<need more info,...I think from Intel>....

April 20, 2000: Intel Corporation. Research discussions.

Brad Anders - Engineering Manager, Knowledge Management

Kristin Short - Systems Engineer, Knowledge Management

Tom Bollinger - Engineering Manager, Manufacturing Training - Media Productions

Todd Houghton - Engineer, Manufacturing Training - Media Productions

John Birchak - Project Manager, Information Technology - Strategy & Technology

May 11, 2000: CMU Graduate School of Industrial Administration, CompFinance visiting students (Employees of KDB, a large Anglo-German investment bank), Demo/presentation.

May 16, 2000: James J. Sutherland, III. Chief, Digital Library Systems. JFCOM/JWFC SPAWAR JTASC. Research discussions.

May 17, 2000: CMU SCS Advisory Board.

May 22, 2000: Clark Zhong's thesis proposal: Apply Multimodal Search and Relevance Feedback. INI.

June 5, 2000: Andrei Villarroel's thesis presentation to IN (Information Networking Institute). Providing Fully-Searchable Video through High-level Scene Understanding. An Automatic Multiple-feature Segmentation and Tracking Algorithm with Change Detection for Surveillance Applications.

June 15, 2000: Dr. Kyu Kim, LTI of Korea. Research discussions.

June 19, 2000: Jerome Yen, Chinese University of Hong Kong. Research discussions.

July 14, 2000: Robert Atkins and Marge Myers, CMU Studio for Creative Inquire. Research/collaboration discussions.

July 20, 2000: Yuri O. Gawdiak, Computer Sciences Division, NASA

July 26, 2000: Carnegie Mellon Research Institute. Presentation/demo on Informedia.

August 9, 2000: Hideaki Tomikawa, Sony-Kihara Research Center, Inc.

August 4, 2000: Yuichi Nakamura, University of Tsukuba. Research discussions.

August 8, 2000: Andy Warhol Museum. Discussion regarding establishing a collaborative effort.

August 10, 2000: Bill Foster, Dreamtime. Discussion regarding working with Informedia technology to index and access NASA's imagery and video resources.

August 15, 2000: General Motors, Ed Schlesinger, James Rillings, Dick Johnson. To discuss Informedia technology.

Faloutsos, C., "Searching, Data Mining and Visualization of Multimedia Data," Invited speaker to Visual Databases VDB5, Fukuoka, Japan, May 10-12, 2000.

Shi, J., "A Framework for Scalable Trainable Image-based Query in Video," DLI2 All-Projects Meeting, Stratford-upon-Avon, England, June 12-13, 2000.

Hauptmann, A., "Auto-Summarization and Visualization Over Multiple Documents and Libraries" DLI2 All-Projects Meeting, Stratford-upon-Avon, England, June 12-13, 2000.

Faloutsos, C., "Internet Topology, Data Mining and Power Laws", HP Labs, Palo Alto, June 2000.

Christel, M., "Development and Evaluation of Digital Video Library Interfaces," presented as a special topic for CMU's Human-Computer Interaction Institute short course on "User Interface Design and Implementation," July 21, 2000.

6 Statement

I certify that to the best of my knowledge (1) the statements herein (excluding scientific hypotheses and scientific opinions) are true and complete, and (2) the text and graphics in this report as well as any accompanying publications or other documents, unless otherwise indicated, are the original work of the signatories or individuals working under their supervision. I understand that the willful provision of false information or

concealing a material fact in this report(s) or any other communication submitted to NSF is a criminal offense (U.S. Code, Title 18, Section 1011).

Principal Investigator Signature: _____