

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

**NSF Cooperative Agreement No. IIS-9817496
Semiannual Progress Report
September 2001**

**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.

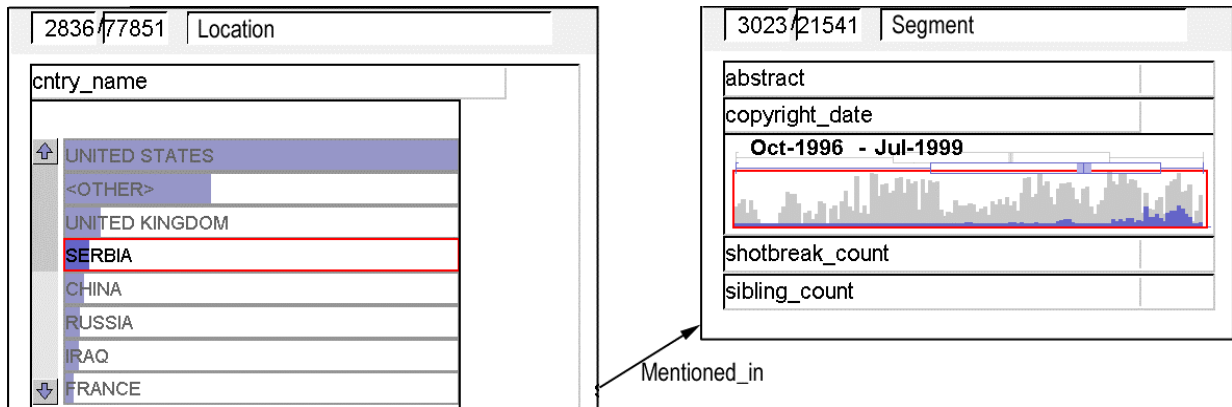
Video information collages are the key technology in Informedia-II and will be built by advancing information visualization research to effectively deal with multiple video documents. A video information collage is a presentation of text, images, audio, and video derived from multiple video sources in 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 **Interactive visualization of digital library data and metadata**

Much current research on digital libraries focuses on named entity extraction and transformation into structured metadata. Examples include entities like events, people, and places, and attributes like birth date or latitude. Unfortunately, this extraction process is not very reliable, and in any case a digital library may contain references to hundreds of thousands of entities. We have been integrating Visage and Visage functionality for visualization and semantic zooming within the Informedia library. Information visualization is a powerful tool for summarizing large sets of structured data about entities and their relationships to uncover overall patterns, and then drilling down into interesting subsets. We have applied this technique to metadata extracted from the Informedia Digital Video Library, and demonstrated examples of conclusions that can be drawn from metadata patterns alone at JCDL'01. Currently, text attributes are handled poorly in terms of both query semantics and interaction speed. Our goal is to overcome these difficulties, so that integrated data and metadata library browsing becomes a continuous interactive activity.



The visual query language allows database-style joins between entities of different types, Dynamic Query filtering of attribute values, visualization of conditional attribute value distributions with histograms, and drill down to individual entities [Derthick1997]. For instance, the distribution of the countries of the geographical locations in the figure above shows the dominance of the US, CNN’s home country, but also concentrations that reflect specific events, as in the case of Serbia. By clicking on histogram bars in the left box, the user can focus on locations in Serbia. This restriction propagates along the *Mentioned_in* relationship to the box on the right. There are 21,541 segments that mention some location, of which 3023 mention a location in Serbia. The dark conditional distribution of the segment *copyright_date* attribute shows that most of the Serbia segments were recorded in 1998-1999, while the light unconditional distribution shows that segments overall are distributed more uniformly from 1996-1999. In some intervals as many as one third of all the news segments mention locations in Serbia.

For more detailed information see Derthick’s JCDL’01 paper.

2.1.2 Hierarchical document clustering of retrieval results

In the last reporting period we investigated the problem of accessing very large search result sets by experimenting with a new text clustering method called Vivisimo (<http://www.vivisimo.com>). Vivisimo is a clustering engine that organizes single or multiple source search results and presents them in a concisely labeled, hierarchical tree-based view, similar to the Windows Explorer style of browsing a file system. We demonstrated the clustering of search results from Informedia’s database at JCDL’01.

The general approach was to develop hierarchical *conceptual clustering* algorithms using ideas from artificial intelligence. The conceptual clustering idea advocates that cluster descriptions should be a primary factor *during the formation of clusters*: good clusters possess good concise descriptions. Thus, a cluster is bad if it does not allow for a good description, even if its internal cohesiveness and external distinctiveness are high according to some distance measure. The basic clustering software is written in C and accepts an input stream of documents in XML format and outputs an XML representation of the hierarchy. The hierarchical folders are labeled with either single words (if they are informative) or multi-word phrases. The C clustering software has a number of parameters, the most used of which is a customizable *stoplist*. If a word is not

informative, it would be placed on a stoplist whose members are words or phrases that can be ignored.

We use a Windows Explorer interaction style for browsing search results, which are presented on a split screen: the left contains the browsable cluster (folder) hierarchy and the right frame contains the subset of search results corresponding to the currently activated folder. The two modes of presenting information complement each other well, according to anecdotal and introspective experience (user studies are underway).

The system (<http://montblanc.se.cs.cmu.edu/informedia.html>) can be run remotely. Currently, most of the processing delay is due to accessing the database, not to the actual clustering, which is fast. For more detailed information, see Valdes-Perez, et.al. JCDL'01 paper from the publication list.

2.1.3 Video mining and classification

We have developed a new tool called VideoGraph for video mining and visualizing the plot structure of a video sequence. The heart of our approach is to group similar shots together, even if they are not consecutive. We use algorithms that “stitch” similar shots into shot-groups, and automatically derive Video Graphs that reveal the logical plot structure of the video clips. This helps us distinguish between different video types, e.g., news stories versus commercials. For example, Figure 1 gives a news clip containing key frames of a shot group and its VideoGraph. Notice the pronounced star-like pattern, which, it turns out, corresponds to an interview. The graph contains only the basic shot-groups, with heavy traffic between the bottom shot-groups (G23, G18, G20, G21), with G23 being the center of the ‘star’-shape. The ‘center’ shot-group G23 is the recurring shot of the interview, which contains shots when the reporter proposed questions to the invited speakers. The shot-groups which interact with G23 heavily, i.e., G18, G20, G21, contains exactly the shots of the invited speakers replying to the reporter’s questions. We also derive features (e.g., number of recurrent shot-groups, etc.) for video mining and classification. Case studies show that good features can be derived automatically for classification of broadcast video into news stories and commercials (91% accuracy), and that Video Graphs reveal the logical structure of video clips. See Pan/Christos JCDL'01 paper for more details.

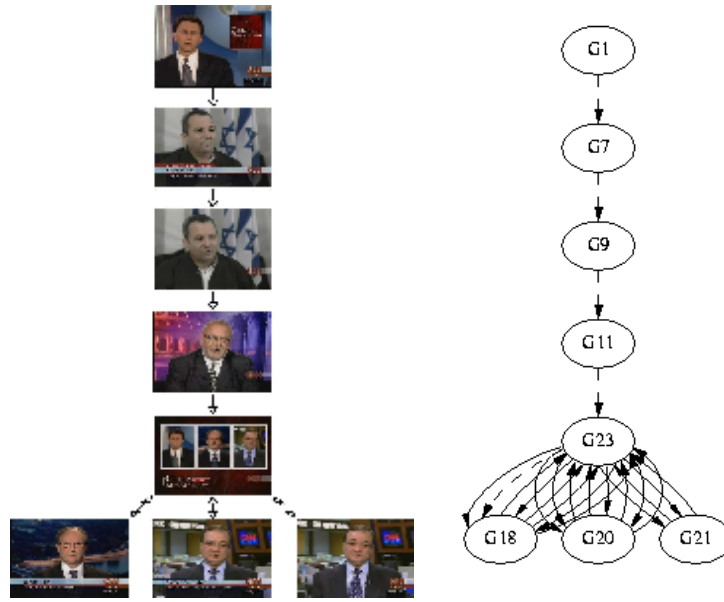


Figure 1. VideoGraph of a news story.

2.2 Information Analysis

2.2.1 Face recognition evaluation

To fully understand the key difficulties in face recognition, we have conducted extensive experiments on a collection of 21,000 images from the CMU/Pitt facial expression (PIE) database, and the AR database from CVC in Barcelona, using Visionics' FaceIt and MIT's Bayesian Eigenface method. We measured to what extent face pose, illumination, facial expression, occlusion, and subject gender affects face recognition performance. We found that facial expression and illumination changes are well-handled, generalizability across face pose (particularly profile view) is poor, and upper face occlusion is much worse than lower face occlusion. Furthermore, we found a surprisingly consistent difference on face recognition rates between genders across 13 test conditionings (based on the AR database which has 60 female and 70 male subjects).

2.2.2 Face recognition using boosted discriminant

Many approaches to face recognition involve the extraction of features from an input image and combining the evidence from these features to discriminate among human subjects. We demonstrated an approach for learning feature combinations in a semi-supervised fashion. Under this framework, the discriminating rule is formed from a weighted sum of the feature evidence for each subject. The features are chosen successively such that at each stage a feature focused on discriminating the objects that remained difficult to distinguish. We formulate the problem for finding the optimal discriminant as an unsupervised learning problem in terms of the pair-wise "confusion" of training subjects in the existing feature space, and in terms of indicator variables that denote the unknown optimal partitions of the training set. We derive an iterative solution that alternates between solving for the discriminant and solving for the indicator variables, each of which is the solution of a corresponding generalized-eigenvalue

problem. We illustrated the encoding of class invariants by learning an illumination-invariant discriminator for faces. This paper is accepted to appear in CVPR 2001 (see Gross/Shi/Cohn CVPR'01 paper). This technique is also being extended to body gait recognition.

2.2.3 Human body and object segmentation in video

To organize visual elements into groups in video depends on various cues presented by the data. Elements often belong in one group because they share a common characteristic feature. For example, images can be partitioned into regions of coherent brightness, texture or motion patterns. Sometimes, however, elements are grouped together not because they are similar to each other, but rather they are all dissimilar to another set of elements. This situation is typified in the perceptual popout phenomena. Completely dissimilar outliers stand out from the background as one group, which is usually considered as the figure that grabs our attention. The differentiation of foreground and background is also closely related to depth segregation. Here the binding of visual elements occurs according to their ordering in the whole organization.

These three types of grouping, i.e., similarity grouping, dissimilarity grouping and figure-ground organization, have often been studied separately. Most existent works focus on similarity grouping, which was advocated over a century ago by the influential school of Gestaltists. Dissimilarity grouping has not been realized in vision science until the 80's. Popout, or salience detection, is not only rarely considered as a grouping problem, but has received little attention in the computer vision community. Though it has long been realized that grouping cues and figure-ground segregation cues should be considered at the same time, the integration of the two has been ad-hoc and sequential. Our work in the last few months has been to unify these three types of groupings into one computational framework, such that image segmentation results from the integration of similarity cues, dissimilarity cues and 3D configuration cues.

We abstract two types of pairwise relationships between visual elements in terms of attraction and repulsion. They encode cues for similarity and dissimilarity grouping respectively. The ordering of visual elements is captured in the asymmetry between pairwise relationships, such that there is a net difference in pairwise repulsion between two elements. Formulated in a graph-partitioning framework, we generalize the normalized cuts criteria to two directed graphs of attraction and repulsion natures. Our model results in Rayleigh quotients on Hermitian matrices, where the real part describes undirected relationships, with positive numbers for attraction, negative numbers for repulsion, and the imaginary part describes the ordering. Globally optimal solutions can be obtained by eigendecomposition. The eigenvectors characterize the optimal partitioning in the complex phase plane, with phase angle separation determining the partitioning of vertices and the relative phase advance indicating the ordering of partitions. We demonstrate the use of our model in detecting either coherent or salient structures, image segmentation together with depth segregation. Within the same framework, we also considered how to integrate prior knowledge seamlessly into data groupings. We consider priors represented by individual group preference of each element, partial *a priori* grouping information and external influence on pairwise comparisons. They are modeled as biases in the grouping process. We incorporate these

biases into graph partitioning criteria. Computationally, this formulation leads to a constrained eigenvalue problem. We demonstrated the effectiveness of this algorithm on image segmentation with user interaction and object detection with spatial attention.

2.2.4 Extracting noun phrases and verbs

The ability to identify nouns and verbs in a textual description gives a better analysis of embedded entities, activities, and the entity-activity interrelationships. We investigated the use of Brill's part-of-speech tagger [Brill1992a, Brill1992b, Brill1993] and University of Pennsylvania's BaseNP Chunker [Ramshaw1995] to extract all possible noun phrases and verbs in textual data. The preliminary experiment applied the original tagging rules to our Informedia News-On-Demand CNN corpus. The rationale was that the original tagging rules were trained on a Wall Street Journal corpus, which is similar to ours in the news dimension. Noun phrase extraction depended on both part-of-speech tagger and BaseNP chunker while verb extraction resulted directly from part-of-speech tagging. In addition, we narrowed the unit of analysis from a news story down to a sentence to obtain the co-occurrence relationship between noun phrases and verbs. Currently, we are evaluating the preliminary results on a subset of CNN news (200 hours). We have planned to use the preliminary tagging results to go through a manual correction effort in order to build an Informedia News-On-Demand corpus for training the part-of-speech tagger and the BaseNP chunker. This research directly gives a mechanism to obtain valid noun phrases for topic assignment and concept association.

2.2.5 Geographic references

We continued exploring the use of several different approaches to named entity tagging for more accurate geographic referencing than our current quick processing baseline, which uses little context adjustment, removes all common and ambiguous terms from the gazetteer match set, and enforces strict term matching. Our initial tagger was too conservative in location assignment, resulting in very few false positives but also overlooking a number of locations mentioned in the source metadata. We are revising this initial tagger to be more aggressive, perhaps augmented with confidence metrics so that the user can direct whether conservative (few false positives, many missed entities) or aggressive (many false positives, few missed entities) strategies are desired for a given information browsing and seeking task.

2.2.6 Shot identification and key frame extraction

We implemented a program that extracts mpeg motion vectors as they are encoded in the mpeg bit stream by standard mpeg encoders. These motion vectors can then be used to identify the camera action (pan, zoom, tilt) and, theoretically, object motion in the video.

2.3 Video Multimodal Query and Retrieval Studies

2.3.1 Participation in NIST TREC 2001

We participated in the video track for NIST TREC 2001. Following is a general outline of the work we performed. Further details of this work and our results will be released in November at the Text Retrieval Conference (TREC).

We entered into our system the video data provided by TREC. We built search and retrieval procedures tailored (especially to image retrieval) to the provided corpus and evaluated it based on the official TREC video track queries.

We integrated into our system two image search engines which are now available on local Informedia clients: WBIIS created by James Wang from Penn State University (<http://www-db.stanford.edu/~wangz/home/>) and a beta version of IBM's QBIC™ (Query by Image Content) system. They are now selections in the local Informedia client only, to be used for further research and investigation. In addition to this, we built an additional probabilistic image search engine that outperforms both WBIIS and QBIC.

We performed an evaluation of face matching systems. We compared Visionics, Eigenfaces, and Eigenflow Based Face Authentication, a system developed at CMU. We found that they were not suited for reliable retrieval for faces in the TREC video data corpus.

With Mediasite, a company founded on early Informedia technology, we integrated and tested speaker identification against the TREC data and found it provided useful results. We also incorporated an improved version of video OCR that gave significantly better results on the video TREC data.

Finally, we used a probabilistic approach for motion vector extraction to help in classifying different camera motions.

2.3.2 Query using face identification

Face identification algorithms have been incorporated into automatic Informedia processing to produce additional descriptive data about the video library. The data includes orientation (frontal, left profile, right profile) and a confidence metric that can be used as a measure of trust when building derivative interfaces from the face metadata.

The user can select which face from the set is the face of interest for subsequent querying, marking the shots with frontal/left profile/right profile marks and the source of the face identification. CMU-developed and commercial processing systems are used to accomplish this.

2.3.3 Automatic title generation

We view the problem of selecting good title words for a document as a variant of an Information Retrieval problem. Each title word is treated as a "document" and selection of appropriate title words as finding relevant "documents". Based on our training collection consisting of 40,000 document and title pairs, we learn the "document" representations for all the title words and apply these learned representations to select appropriate title words over 10,000 test documents. Compared to other learning

approaches, namely K nearest neighbor approach, a Naïve Bayesian approach and a variant of a machine translation model, we find that our approach is significantly better as indicated by the F1 metric. For more detailed information see Jin/Hauptmann ICML 2001 paper.

We implemented several statistical title generation methods using a training set of 21190 news stories and evaluated them on an independent test corpus of 1006 broadcast news documents, comparing the resulting titles based on manual transcription to the titles from automatically recognized speech. Both F1 and the average number of correct title words in the correct order were used as evaluation metrics. The 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. For more detailed information see Jin/Hauptmann HLT 2001 paper.

We also compared automatically generated titles for machine-translated documents using several different statistics-based methods. A Naïve Bayesian, a K-Nearest Neighbor, a TF-IDF and an iterative Expectation-Maximization for title generation were applied to 1000 original English news documents and again to the same documents translated from English into Portuguese, French or German and back to English using SYSTRAN. The Auto Summarization function of Microsoft Word was used as a base line. Results on several metrics show that the statistics-based methods of title generation for machine-translated documents are fairly language independent and title generation is possible at a level approaching the accuracy of titles generated for the original English documents. For more detailed information see Jin/Hauptmann IJCAI 2001 paper.

2.3.4 Meta-scoring: Automatically evaluating term weighting schemes in IR without precision-recall

The central task of an information retrieval system is to find documents relevant to a given query and rank them in order of decreasing match with the query statement. This implies the calculation of some quantitative measurement of similarity between the query and each document. The traditional evaluation method is through precision and recall, which requires human judgments about the relevance of the documents to the queries. However, collecting human relevance judgments for every document to every query is costly and time consuming, plus human judgment is highly subjective.

We developed a method that can automatically evaluate performance of different term weighting schemes in information retrieval without resorting to precision-recall based on human relevance judgments. Specifically, the problem is: given two document-term matrices generated from two different term weighting schemes, can we tell which term weighting scheme will perform better than the other? We developed a meta-scoring function, which takes as input the document-term matrix generated by some term weighting scheme and computes a “goodness” score from the document-term matrix. In our experiments, we found that this score is highly correlated with the precision-recall measurement for all the collections and term weighting schema we tried. Thus, we conclude that our meta-scoring function can be a substitute for the precision-recall measurement that needs relevance judgments of human subjects. Furthermore, this meta-scoring function is not limited only to text information retrieval can be applied to fields

such as image and DNA retrieval. For more detailed information, Jin/Faloutsos/Hauptmann ACM SIGIR 2001 paper.

2.3.5 Web formats promoting flexible presentations of video metadata

Based on prior success within the Informedia Project regarding the automatic generation of video metadata and interfaces to deal with the volume of material from growing video repositories and expanded metadata coverage, Christel was invited to participate in a “Reforging the Links” Workshop at the University of Wisconsin Sept. 10-11, funded by the Television Future Fund of the Corporation for Public Broadcasting (<http://reforginglinks.uwex.edu/>). This workshop discussed the general theme of links between public television and the university in the digital age, with emphasis on digital video and digital asset management. One of the topics explored at the workshop dealt with web presentations of video metadata, tailorable by the user through XSLT.

3 Notable Outreach and Inclusion Activities

- Mike Christel presented a paper “Accessing News Video Libraries through Dynamic Information Extraction, Summarization, and Visualization” at the JCDL workshop “Visual Interfaces to Digital Libraries - Its Past, Present, and Future” held on June 28, 2001 (see <http://vw.indiana.edu/visual01/jcdl.html> for details, PDF of the paper, and slides). The primary aim of the workshop was to raise the awareness of several interconnected fields of research related to the design and use of visual interfaces to digital libraries, especially in information visualization, human-computer interaction, and cognitive psychology. This workshop also aimed to stimulate participants to reflect on the state of the art in their own fields by identifying challenging issues concerning visual interfaces and thereby fostering a multidisciplinary research agenda for future research and development.
- We collaborated with the Indiana Center for Database Systems, in particular the Indiana Telemedicine Incubator (ITI) Project’s digital video system for medical education. This EduMed system has a repository of segmented and annotated medical videos which first and second year medical school faculty and students can browse, retrieve and play. We generated metadata automatically through Informedia processing for five representative videos from this collection, and shared that with the ITI staff. The ITI staff was primarily interested in Informedia speech recognition and text metadata derivation for content-based indexing, along with segmentation issues. This collaboration helped the ITI staff refine their prototype and define their operational requirements for video processing in the medical education domain.
- We demonstrated the Informedia system to public broadcasters and university extension programs, discussing the automatic derivation of metadata for video enabling full-content search and retrieval, and the special interface needs for dealing efficiently and effectively with growing amounts of metadata for video collections that are growing as well. Michael Christel was an invited panelist to the “Reforging the Links” Workshop at the University of Wisconsin Sept. 10-11, funded by the Television Future Fund of the Corporation for Public Broadcasting (<http://reforginglinks.uwex.edu/>). This workshop discussed the general theme of

links between public television and the university in the digital age, with emphasis on digital video and digital asset management.

- Based on Informedia DLI-2 work and research directions, we will again update and deliver “Multimedia”, course 20-791, as part of the Carnegie Mellon E-Commerce Master's Program (<http://www.ecom.cmu.edu>), the first E-Commerce degree program in the country. Course Description: Until recent years, most computing tasks dealt with numerical, text, and symbolic data, and Computer Science has emphasized these data types. Digital representations of audio, video, and images are now becoming quite common. With the advent of relatively cheap, large online storage capacities, network transmission speeds and advances in digital compression, comprehensive sources of multiple media (Text, Image, Video and Audio) can be easily stored and made available. Collecting and intelligently integrating these multiple media opens up opportunities for novel business applications. Consequently, an understanding of multimedia is essential for many e-commerce businesses.

This course teaches students to work with multiple media on computers. Students learn the issues involved to capture, process, compress, search, index, store, and retrieve various kinds of continuous media. Projects require work with audio, scanned images, digital video, and other media, all in digital form. Readings and lectures provide a conceptual and technical framework for multimedia work. After completing this course, students will be able to:

- Appreciate the role of multimedia in E-Commerce
 - Understand the concepts underlying multimedia creation, representation and transmission
 - Create media for the Web using various software tools to manipulate audio, images and video
- Informedia was mentioned in the article “Upstream: Video Searching”, by David Voss, (<http://www.techreview.com/magazine/jul01/upstream.asp>) which appeared in MIT's *Technology Review*, July/August 2001 issue. This article discusses the area of searching and retrieving in video libraries:

At Carnegie Mellon University, researchers are creating a digital library that combines natural-language processing, speech recognition and image analysis. “The integration of these different technologies is the key,” says Howard D. Wactlar, director of the Informedia Digital Video Library Project at Carnegie Mellon. A prototype captures news broadcasts from around the world and stores them, along with summaries or storyboards. Someone can then type in a question, or just utter the question aloud: “Tell me about oxygen problems on the Russian space station Mir.” All the relevant news clips are displayed as frame icons you can click on. The system is also incorporating face recognition to make it possible to call up all the clips of a particular person.

- We continue our partnership with the European Commission Information Societies Technology (IST) program sponsored by European Chronicles On-line (ECHO). The

main objectives of the project are to develop a long-term reusable software infrastructure to support digital film archives, to provide web-based access to collections of historical documentary films of great international value, and to increase the productivity and cost effectiveness of producing digital film archives. The project will develop and demonstrate an open architecture approach to distributed digital film archive services. The open architecture will support service extensibility and interoperability. The distinct features of the ECHO system will be: semi-automatic metadata extraction and acquisition from digital film information, non-English speech recognizers (Italian, French, Dutch, German) for the purpose of indexing, searching and retrieval, cross-language retrieval capabilities, intelligent access to digital films, automatic film summary creation, collection mechanisms, and privacy and billing mechanisms.

- We continue our cooperative research and technology transfer relationship with three Chinese research and educational institutions in order to further extend the system's use, increase its capabilities, and enrich its distributed corpus with multilingual Chinese video content. The institutions involved are The Chinese University of Hong Kong (PI's: Jerome Yen, Joseph Hui, Michael Lyu), the South China University of Technology in Guang Zhou (PI: Ling Zhang), and the Academia Sinica in Taiwan (PI: Der-Tsai. Lee). We have transferred an independently operating demonstration library with seed content to Hong Kong and established specific research relationships amongst individual researchers at the cooperating institutions to replicate and extend the underlying video analysis and extraction techniques in both Mandarin and Cantonese variants.
- Our participation in TalkBank continues. TalkBank is a multimedia database of communicative interactions, that 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.
- We continue our ongoing collaboration with General Motors Research Lab. They are 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 are ongoing and still in their initial stage, but have yielded interesting development ideas and plans to collaborate.
- We continued cooperation with the Open Video Project at UNC Chapel Hill (<http://www.open-video.org>) by providing segmented MPEG-1 public domain video to their open source library. The purpose of the Open Video Project is to collect and make available a repository of digitized video content for the digital video, multimedia retrieval, digital library, and other research communities. Researchers can use the video to study a wide range of problems, such as tests of algorithms for automatic segmentation, summarization, and creation of surrogates that describe video content; the development of face recognition algorithms; or creating and evaluating interfaces that display result sets from multimedia queries. Because researchers attempting to solve similar problems will have access to the same video

content, the repository is also intended as a test collection that will enable systems to be compared, similar to the way the TREC conferences are used for text retrieval. This repository is hosted as one of the first channels of the Internet 2 Distributed Storage Infrastructure Initiative (<http://dsi.internet2.edu/>), a project that supports distributed repository hosting for research and education in the Internet 2 community.

4 Technology Transfer

Nothing new to report

5 Journal and Conference Proceeding Publications

- Bi, Z., Faloutsos, C., Korn, F. "The "DGX" Distribution for Mining Massive, Skewed Data", The Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2001), San Francisco, CA, August 26-29, 2001. ("Best Paper Runner-Up" Award.)
- Christel, M., Wactlar, H., Hauptmann, A. "Improving Access to Digital Video Archives through Informedia Technology", Journal of the Audio Engineering Society, 49(7/8), July/August 2001.
- Christel, M., Maher, B., Begun, A. "XSLT for Tailored Access to a Digital Video Library", Joint Conference on Digital Libraries (JCDL 2001), Roanoke, VA, pp. 290-99, June 24-28, 2001.
- Christel, M., Warmack, A. "The Effect of Text in Storyboards for Video Navigation", IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Salt Lake City, UT, May 7-11, 2001.
- Christel, M., Huang, C. "SVG for Navigating Digital News Video", To appear in ACM Multimedia 2001, Ottawa, Canada, September 30-October 5, 2001.
- Derthick, M. "Interactive Visualization of Digital Library Data and Metadata", Joint Conference on Digital Libraries (JCDL'01), Roanoke, VA, p. 453, June 24-28, 2001 (presented as a demo).
- Georgiadis, C., Triantafillou, P., Faloutsos, C. "Fundamentals of Scheduling and Performance of Video Tape Libraries", Multimedia Tools and Applications Journal (MTAP) 2001.
- Gross, R., Shi, J., Cohn, J. "Where to go with Face Recognition", To appear in Third Workshop on Empirical Evaluation Methods in Computer Vision, IEEE Conference on Computer Vision and Pattern Recognition 2001 (CVPR'01), Hawaii, December 9-14, 2001.
- Jin, R., Hauptmann, A., Faloutsos, C. "Meta-scoring: Automatically Evaluating Term Weighting Schema in IR without Precision-Recall", The 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, Louisiana, September 9-13, 2001.

- Jin, R., Hauptmann, A. "Title Generation for Machine-Translated Documents", Seventh International Joint Conference on Artificial Intelligence (IJCAI '01), Seattle, WA, August 4-10, 2001.
- Jin, R., Hauptmann, A. "Learning to Select Good Title Words: A New Approach based on Reversed Information Retrieval", International Conference on Machine Learning (ICML '01), Williams College, Berkshires, MD, June 28- July 1, 2001.
- Jin, R., Hauptmann, A. "Automatic Title Generation for Spoken News", Human Language Technology Conference (HLT'01), San Diego, CA, March 18-21, 2001.
- Korn, F., Pagel, B., Faloutsos, C. "On the 'Dimensionality Curse' and the 'Self-Similarity Blessing'", IEEE TKDE Transactions on Knowledge and Data Engineering (TKDE), special issue (best of ICDE2000), 13(1), pp. 96-111, January/February 2001.
- Mahamud, S., Hebert, M., Shi, J. "Object Recognition using Boosted Discriminants", , To appear in IEEE Conference on Computer Vision and Pattern Recognition 2001 (CVPR'01), Hawaii, December 9-14, 2001.
- Montgomery, A.L., Faloutsos, C. "Identifying Web Browsing Trends and Patterns", IEEE Computer, 34(7), pp. 94-95, July 2001.
- Palmer, C., Pesenti, J., Valdes-Perez, R., Christel, M., Hauptmann, A., Ng, D., Wactlar, H. "Hierarchical Document Clustering of Digital Library Retrieval Results", demonstration at JCDL 2001 (Roanoke, VA, June 2001).
- Pan, Jia-Yu, Faloutsos, C. "VideoGraph: A New Tool for Video Mining and Classification", Joint Conference on Digital Libraries (JCDL 2001), Roanoke, VA, p. 116-117, June 24-28, 2001.
- Proietti, G., Faloutsos, C. "Accurate Modeling of Region Data", IEEE Transactions on Knowledge and Data Engineering (TKDE) (to appear).
- Ridel, E., Faloutsos, C., Gibson, G., Nagle, D. "Active Disks for Large-Scale Data Processing", IEEE Computer 34(6), pp. 68-74, June 2001.
- Santos, R.F., Filho, Traina, A.J.M., Traina, C.Jr., Faloutsos, C. "Similarity Search without Tears: The OMNI Family of All-purpose Access Methods," Proc. International Conference on Data Engineering (ICDE), Heidelberg, Germany, April 2-6, 2001, pp. 623-630.
- Traina, C., Traina, A., Faloutsos, C., Seeger, B. "Fast Indexing and Visualization of Metric Datasets using Slim-trees", IEEE TKDE (Transactions on Knowledge and Data Engineering) (to appear).
- Traina, A., Traina, C., Papadimitriou, S., Faloutsos, C. "Tri-Plots: Scalable Tools for Multidimensional Data Mining", The Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2001), San Francisco, CA, August 26-19, 2001.

- Valdes-Perez, R.E., Wactlar, H., Pesenti, J., Palmer, C., Ng, D., Hauptmann, A., Christel, M. "Demonstration of Hierarchical Document Clustering of Digital Library Retrieval Results", Joint Conference on Digital Libraries (JDCL'01), Roanoke, VA p. 451, June 24-28, 2001 (presented as a demonstration).
- Wang, M., Madhyastha, T., Change, N.H., Papadimitriou, S., Faloutsos, C. "Data Mining Meets Performance Evaluation: Fast Algorithms for Modeling Bursty Traffic", Submitted to the IEEE 18th International Conference on Data Engineering (ICDE '02), San Jose, CA, February 26-March 1, 2002.
- Yu, S.X., Shi, J. "Segmentation with Pairwise Attraction and Repulsion", International Conference on Computer Vision 2001, July 9-12, 2001, Vancouver, British Columbia, Canada (http://www.cs.cmu.edu/~jshi/jshi_publication.htm).
- Yu, S.X., Shi, J. "Grouping with Directed Relationships", Third International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR'01), September, 2001.
- Yu, S., Shi, J. "Grouping with Bias", To appear in the proceedings of Neural Information Processing Systems (NIPS'01), Vancouver, British Columbia, Canada, December 3-8, 2001.
- Yu, S., Shi, J. "Understanding Popout: Pre-attentive Segmentation through Nondirectional Repulsion", To appear in IEEE Conference on Computer Vision and Pattern Recognition 2001 (CVPR'01), Hawaii, December 9-14, 2001.

6 Presentations, Demonstrations, and Industry Visitors

- September 2001 Faloutsos: "Searching and Data Mining in Multimedia Databases", Distinguished Lecture Series, University of Southern California
- August 20, 2001 Wactlar: Presentation to Chinese University visit to discuss digital library research.
- Prof. PAN Yunhe (Head of the Delegation), President of Zhejiang University, Member of Chinese Academy of Engineering
- Dr. GAO Wen (Deputy Head of the Delegation (Chinese PI of the One Million DL Project). Deputy President of Graduate School, Chinese Academy of Science
- Prof. CHI Huisheng, Vice President of Beijing University
- Prof. HU Dongcheng, Vice President of Tsinghua University
- Prof. XU Zhong, Vice President of Fudan University
- Prof. ZHANG Yibin, Vice President of Nanjing University

- Mr.Chen Jianping, Vice director, Department of Social Development,
State Planning Commission of China
- Mr. GUO Xinli, Vice General Director, “211 Project” Office,
Ministry of Education of China
- July 9-13, 2001 Wactlar: Lecture at DELOS International Summer School of Digital
Library Technologies, ISDL 2001, Pisa, Italy.
- July 5, 2001 Christel: Presentation to Andrew’s Leap, a six-week advanced
summer program for high school students 13-17 years of age,
sponsored by CMU’s School of Computer Science.
- June 27-28, 2001 Wactlar/Christel/Hauptmann: DLI2 PI Meeting, Roanoke, Virginia
- .June 13-15, 2001 Wactlar: Presentation at DELOS Brainstorming meeting on “Digital
Libraries: Future Research Directions for a European Research
Programme”, Dolomites, Italy. Sponsored by The DELOS Network
of Excellence on Digital Libraries, in cooperation with the Cultural
Heritage Applications Unit of the 5th FP IST. (USC), Los Angeles,
CA.
- May 10, 2001 Wactlar: Meeting with Mass Mutual to discuss ongoing research.
- May 13-16, 2001 Wactlar: ECHO meeting at Techmath,Kaiserslautern, Germany.
- May 29-31, 2001 Wactlar: “Multi-Document Summarization and Visualization in the
Informedia Digital Video Library”. Presented at The 12th
International Conference on New Information Technology, Global
Digital Library Development in the New Millenium (NIT’01),
Tsinghua University, Beijing, China.
<http://web.simmons.edu/~chen/nit/>.
- April 30, 2001 Christel: Informedia demo and discussion with Hitachi:

Dr. Hiromichi Fujisawa, Senior Chief Researcher, Hitachi Central
Research Laboratory, Tokyo, Japan

Tetsuro Kiyomatsu, Chief Engineer, Hitachi Mechatronics Systems
Division, Tokyo, Japan

Tetsuo Machida, General Manager, Marketing and Planning
Operation, Hitachi Mechatronics Systems Division, Tokyo, Japan
- June 19, 2001 Wactlar/Christel: Demo and discussion with NTT:

Toshiyuki Kurahashi, Assistant Manager, Broadband Project, Audio
Visual Production Division, NTT Learning Systems Corporation

Akira Mino, Executive Manager, Broadband Project, Audio Visual
Production Division, NTT Learning Systems Corporation

Bill Hayes, Systems Analysis & Software Design, Japanese/English
Translation

Sukeyuki Arima, President and CEOP, MediaSite, KK, Japan

- Steven R. (Randy) Olson, Vice President, Software Engineering, MediaSite.
- April 27, 2001 Ng: Visit from Embratel, the WorldCom/MCI branch in Brazil to discuss Informedia and their ongoing Multimedia Digital Library. Tulio Gontijo Rocha, Network Projects Coordinator.
- April 23, 2001 Wactlar: Interview with David Voss, Technology Review, for article: "Upstream: Video Searching", Technology Review, July/August 2001
<http://www.technologyreview.com/magazine/jul01/upstream.asp>
- April 3, 2001 Christel: Visit and discussion with Masato Kawade, Manager, Human Media Promotion Group, Information Technology Research Center, Omron <http://www.omron.co.jp>.
- March 26, 2001 Wactlar: Presentation to Western Psychiatric Institute and Clinic regarding possible research collaboration.
- March 19, 2001 Faloutsos: "Searching and Data Mining in Multimedia Databases", Invited Distinguished Lecture, University of Alberta, Canada.
- March 2, 2001 Hauptmann: Presentation on Informedia at INI Research Seminar.
- Feb 7, 2001 Hauptmann: Informedia presentation to visitors from Telefonica (Spain).

7 Other Report References

- [Ramshaw1995] Ramshaw, Lance and Marcus, Mitchell "Text Chunking Using Transformation-Based Learning", *Proceedings of the Third ACL Workshop on Very Large Corpora*, MIT, June, 1995. (<http://xxx.lanl.gov/e-print/cmp-lg/9505040>)
- [Brill1992A] Brill, E. and Marcus, M. "Tagging an Unfamiliar Text With Minimal Human Supervision", *Proceedings of the Fall Symposium on Probabilistic Approaches to Natural Language*, American Association for Artificial Intelligence (AAAI), 1992.
- [Brill1993] Brill, E. "A Corpus-Based Approach to Language Learning", Ph.D. Thesis, Department of Computer and Information Science, University of Pennsylvania, 1993.
- [Brill1992B] Brill, E. "A Simple Rule-Based Part of Speech Tagger", *Proceedings of the Third Conference on Applied Natural Language Processing*, ACL, Trento, Italy, 1992.
- [Derthick1997] An interactive visual query environment for exploring data. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '97)*, pages 189-198, 1997.