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

NSF Cooperative Agreement No. IIS-9817496
Semiannual Progress Report
March 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 Contextual term phrases in video documents

The work of analyzing contextual term phrases from last reporting period has shifted its focus on defining contextually coherent news segments (news stories) from broadcasting news programs. Through observation and experiment [Ng2000, Zhong2000], we obtained further data on the use of contextual term phrases to extend the summarization and presentation dimensions for broadcast news 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. Our work showed that for broadcast news, a sequential segment of video may contain a detailed news story as well as news headlines for stories shown later in the same news program. These forward references to other stories need to be filtered out in order to derive better contextual term phrases for the news story in focus. The current work of integrating contextual information with temporal and syntactical evidences will benefit the summarization research. We will continue the work in finding contextual term phrases and their associations in the coming period.

2.1.2 Adaptive query-based and adjustable storyboards

We published the results of our storyboard experiments focusing on the utility of transcript text in storyboards for news video navigation in a paper accepted for publication at ICASSP 2001 (see publications list). A storyboard, also referred to as

filmstrip, is a time-ordered series of images drawn directly from the video, presented all at once, typically scaled down in size. As noted in our prior report, we developed a series of storyboard interfaces with added transcript text features that were used in a controlled experiment focusing on the utility of transcript text in storyboards for news video navigation. We explored the role of text synchronization and reduction within storyboards. We explored whether such text features resulted in improvements in video navigation, and, if so, whether the amount of text and its synchronization with video imagery affected the navigation task.

We found that storyboard surrogates clearly improved with the addition of text, and participants favored interleaved text/image presentation, where transcript text is compressed to a single line or a few lines of caption for each row of images. Interleaved treatments were voted the top two storyboard schemes (see Table 1). Despite users' preference to have all the text available, they were able to accomplish the information finding task in less time with the interleaved reduced text format (see Figure 1).

Results	**NoText** (images w/ no text)	**AllByRow** (images w/ full text by row)	**BriefByRow** (images w/ brief text by row)	**All** (images above full text)	**Brief** (images above brief text)
Mean completion time in seconds	192	160	117	137	162
Mean ranking for treatments (1=favorite, 5=least favorite)	4.60	1.12	2.52	3.04	3.72

Table 1: Mean completion time and mean ranking for treatments

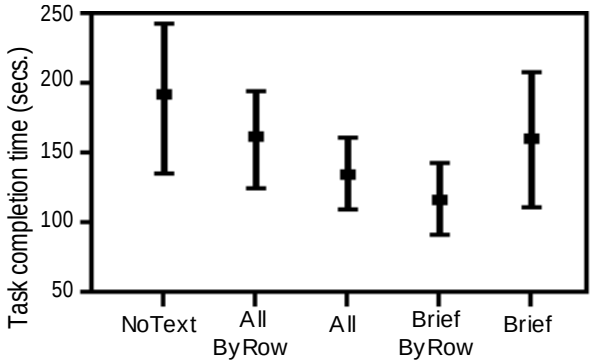

Figure 1. Task times with 95% confidence intervals

These results have implications for video collages. If interleaving is done in conjunction with text reduction, to better preserve and represent the time association between lines of text, imagery and their affiliated video sequence, then storyboards with great utility for information assessment and navigation can be constructed. Such combined transcript sequencing and compression could become part of the feature set offered by next generation digital video players. These conclusions are based on surrogates for a fact-finding task against a particular genre of video: news. Video preview features that work well for one genre may not be suitable for a different type of video: a recent study found that shot images were used most frequently and rated most useful for news, travel, and sports, with lowest ratings and least use for classroom lecture and conference presentations [Li2000] (this study did not address transcript text). Future work includes examining the utility of storyboards with text for open-ended browsing tasks, and with other genres where the visual content may not be as rich and varied.

2.1.3 Dynamic panorama generation

Another approach to collapsing a video sequence across time is to present it as a single still image with greater resolution than the originating video. Such a technique works well with camera pans, where a landscape setting or participants in a given area can be shown all at once to the viewer. This panoramic technique reduces a large amount of redundant information in a video segment based on the existence of large amount of overlapping image information. Dynamic panorama generation has been created and integrated into the Informedia library interface, whereby a user can select a video time range and have a single panoramic image automatically generated in real-time where redundant visual information is collapsed appropriately (see publications list, Villarroel Master's Thesis). Panoramas have been tested for camera pans across both horizontal and vertical dimensions.

The panorama generator may be integrated into automated video processing for enhanced metadata creation. For example, camera pans can be identified automatically with high degrees of confidence. Other processing techniques might detect features such as overlaid text with durations across the camera pan but changing locations, such as a large sign that is panned across in time. By detecting overlaid text against a generated panorama image that shows the whole sign, the resulting text will be more complete and accurate.

2.2 Information Analysis and Semantic Summarization

2.2.1 Face identification

Based on the PhD thesis and continuing post-doctoral work of Henry Schneiderman [Schneiderman2000a, 2000b] face identification has been greatly enhanced over the previously used Rowley algorithm [Rowley1998]. In pilot evaluations comparing Schneiderman's methods with our previous methods and that of commercial software Visionics, we found Schneiderman's method resulted in 80% of faces correctly identified for a candidate news broadcast, with Visionics FaceIT software and our previous methods producing less than 60% accuracy.

Face identification algorithms will be 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 identified faces will be incorporated in future summarization techniques showing key persons across time and space for video collections defined interactively through queries and refinement.

2.2.2 Enhanced speech recognition

Sphinx 3.3 is now running under the Microsoft Windows NT/2000 operating systems and uses filter graphs to better address the audio portion of multimedia files in multiple formats beyond just MPEG-1. Sphinx 3.3 produces significantly lower error rates on broadcast video narrative compared to the earlier Sphinx 2 processing. The better recognizer allows more accurate text transcripts to be passed to additional Informedia

processing such as named entity extraction for finding geographical references, as well as better timing information aligning spoken dialogue with the video. The enhanced metadata, along with improved alignment, are building blocks for producing collages that summarize multiple video sequences.

2.2.3 Automatically generating meaningful titles

We published the results of our automatic title generation studies in a paper presented at the Human Language Technologies Conference in March 2001 (see publications list). We implemented several statistical title generation methods (NBL, NBF, TF.IDF, KNN and EM) using a training set of 21,190 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. We use both an F1 metric and the average number of correct title words in the correct order as evaluation metrics. We conclude that KNN is desirable for title generation especially when overlap in content between training dataset and test collection is large. The fact that NBL outperforms NBF and TF.IDF out performs NBL suggests that we need to distinguish important document words from those trivial, unimportant words. Finally, we are encouraged that title generation from speech-recognized documents is possible at a level approaching title generation from perfect text transcription.

2.2.4 Shot identification and key frame extraction

Video is commonly broken down into sets of contiguous shots. Various video content analysis research groups and video libraries use this shot decomposition, with automation techniques surveyed in a recent paper [Gargi2000]. The automated procedure for producing shots in the Informedia library has been improved by using adaptive local difference thresholds of successive color histograms to select new shots, rather than fixed color histogram differences, with testing and evaluation being done as a participant in the NIST TREC-2001 Video Retrieval Track (http://www-nlpir.nist.gov/projects/t01v/). We added extensions to produce VideoGraphs, new tools for video mining and visualizing the structure of a video sequence. Case studies show that good features can be automatically derived for classification of broadcast video into news stories and commercials (91% accuracy), and that VideoGraphs reveal the logical structure of video clips. For more details, refer to the JCDL 2001 paper by Pan and Faloutsos. Future improvements to selecting a representative image, or key frame, for each shot will include confidence metrics from face identification, identification of text in the image, and heuristics suggested by participation in TREC-2001.

2.2.5 Context analysis: concept association

In order to achieve the technology of *video information collages*, we have re-explored our digital video library (DVL) to garner extra-analyzed results from existing analysis and processing applied on our video data. As our Informedia research progresses toward adaptive visualization and summarization over a set of videos, the availability and extensiveness of both quantitative and qualitative features in each news story become vital for analytic computation and meaningful presentation of a large set of video data. Some examples of features being analyzed but not yet made available from our library

are: face/non-face images, number of faces in an image, and text/non-text images. In conjunction to such rediscovering, we are actively adding and upgrading our DVL's analytic modules with our latest research results.

We have implemented named entity extraction along the lines of [Kubala1998] to better classify transcript text as places, people, and organizations, according to the research plan outlined in [Christel2000]. Such additional text processing is necessary to disambiguate place names to enable geo-collages and other summarization techniques. For example, the proper noun "Washington" could refer to a western state in the United States, its capital city, or could be a person's name. Contextual cues such as "Seattle, Washington", "Washington, D.C." and "since the time of George Washington" distinguish these different meanings. Through contextual analysis on the source metadata, the system will be able to better classify proper nouns as persons' names or places, and more accurately assign location information.

Hidden Markov models (HMMs) can be used to achieve this level of analysis. HMMs have proven effective for automatically tagging entities, including locations, in text output from speech recognizers, where such text lacks punctuation cues [Kubala1998]. Alternative approaches are either rule-based [Aberdeen1995] or language model based [Yoshihiko1999]. We are currently 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. Further studies will be conducted to determine the accuracy and benefits associated with the additional processing.

2.2.6 Learning spectral image segmentation in video

Good image segmentation of a video is essential for recognizing and indexing objects in video. We are continuing our research in two directions. First, we have deepened our understanding of the grouping and segmentation process in a statistical framework, presenting 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 (refer to Meila and Shi papers for details). 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). We are studying how more complex shape information can be encoded in this framework.

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

2.3.1 Face matching, clustering, and naming

Via improved face identification, as discussed in Section 2.2.1, many more images are marked as containing faces. Some of the tagged faces do not have enough visual detail to support recognizing and resolving the face to a particular person or for use in face matching. For example, profiles can be recognized, but a profile of person *X* will not

match a frontal view of person *X* using traditional image-only techniques. We are exploring the use of the co-occurrence of text features like spoken dialogue with face identification to associate certain text with a cluster of faces. We anticipate that the ability to label faces, as discussed in "Name-It" Informedia reports [Satoh1997, Houghton99], will be improved through both face confidence metrics and face orientation information. We will continue collaborating with Shin'ichi Satoh, National Institute of Informatics, Japan (http://www.rd.nacsis.ac.jp/~satoh/) on this effort, which enables faces to be added to timelines, maps, and other exploratory interfaces to show key people during a specified time period, location, or other criteria.

2.3.2 Text and image clustering

As the contents of digital libraries grow, accessing their contents will become as unwieldy as accessing the World Wide Web is now. In response to a user's typical information need, a system will output hundreds of matches which no user will have the inclination to examine in detail. We have investigated this problem of accessing very large search results 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 believe that clustering of search results is a good approach in the face of insufficient information about user intents. The idea is that a moderately large mass of search results (perhaps 100-1000) is presented as an easily navigable hierarchy, with the user relied upon to act on his or her intent by selecting to expand one or more clusters after viewing their descriptions. This work will be presented by Palmer et al. as a demonstration at JCDL 2001 (see publication list). Future work includes integrating the newest algorithms into the Informedia server, exploring extensions for image clustering, and conducting user studies.

2.3.3 Semantic zooming

We have built exploratory map interfaces adjusting level of detail. Geographic metadata for video, coupled with gazetteer information concerning a hierarchy of city-state/province-country-continent relationships, provides for semantic zooming over multiple video stories. We have created interfaces that allow stories to be summarized across either or both of state/province views and country views. Coupled with other information visualization techniques such as dynamic query sliders, the maps can be animated under user control to show geographic relationships across time and topics.

Visage is a data exploration and visualization system developed by Maya Design Group and Carnegie Mellon University [Roth97]. We have been working on integrating Visage and Visage functionality for visualization and semantic zooming within the Informedia library; Wactlar and Derthick will demonstrate this work and its potential at *JCDL 2001*. Through "brushing" [Becker1987], a form of selection provided in Visage, coordination across visualizations can be provided, e.g., operations on maps are reflected on timelines and vice versa.

2.3.4 Tailorable interfaces accounting for automatically derived metadata

Automated video processing is errorful, but integrating across techniques can increase confidence in particular metadata. For example, a face identified by multiple techniques can be relied on in summary interfaces with greater trust than a face identified with low confidence by only a single technique. Video analysis algorithms have been revised to preserve confidence metrics where possible in the Informedia database. The user is then presented with greater control as to the level of trust required in producing an interface.

2.3.5 Web formats promoting flexible presentations of video metadata

The Informedia Project is migrating to an XML/XSLT delivery infrastructure, where users via web browsers will be able to query the library and retrieve data in XML format corresponding to the metadata associated with video in the library. Depending on user locale and the rights associated with the video clips, the video itself may also be accessible. The architecture of the Informedia library system is being re-engineered, migrating from a two-tier client-database server system to a multi-tier client: user interface, presentation transformations (e.g., XSL to convert XML into HTML presentation for web-based client), video library logic (e.g., ideal use of shot images within slide shows and synchronized presentations containing text fragments), and database.

Scalable Vector Graphics (SVG) [refer to http://www.w3.org/Graphics/SVG] is a language for describing two-dimensional graphics in XML. Map and "visualization by example (VIBE)" interfaces for video exploration have been built as SVG, generated from the same XML data representing a set of video from the Informedia library. The use of XML, XSL, XPath, SVG and other W3C recommendations allows client browsers to access the Informedia library and present multiple summaries of the video. Through client-side transformations of XML into HTML, SVG, or other formats, a user can manipulate summaries, i.e., video collages, dynamically to explore features of interest specific to his or her needs, without burdening the Web server or requiring further download time overhead. See the Christel, Maher and Begun JCDL paper for details.

2.3.6 Reducing spatial datasets through fractal dimensions

When managing the increasing volume of data generated by organizations, a frequently arising question is "what part of this data is really relevant to be kept?" Notice that usually the relations of the database have many attributes that are correlated with the others. Attribute selection is a classic goal, as well as battling the "dimensionality curse" [Berchtold1998, Pagel2000]. Dimensionality curse and dimensionality reduction are two issues that have retained high interest, for machine learning, multimedia indexing, clustering, and data mining. A careful chosen subset of attributes improves the performance and efficacy of a variety of algorithms. This is particularly true with redundant data, as many datasets can largely be well approximated in fewer dimensions. This can also be seen as a way to compress data, as only the attributes that maintain the essential characteristics of the data are kept [Fayyad98].

We have developed a fast, scalable algorithm, Fractal Dimension Reduction (FDR), to quickly select the most important attributes (dimensions) for a given set of n-dimensional vectors. In contrast to older methods, our method has following desirable properties: (a) it

does not do rotation of attributes, thus leading to easy interpretation of the resulting attributes; (b) it can spot attributes that have non-linear correlations; (c) it requires a constant number of linear passes over the dataset; and (d) it gives a good estimate on how many attributes we should keep. The idea is to use the intrinsic or fractal dimension of the given dataset, and to drop attributes that do not affect it. We applied our method on real and synthetic datasets, where it gave fast and good results. For example, when applied to 12,000 face vectors from the Informedia video library, FDR found that 5 of the original 16 vectors were enough to characterize the data set. Instead of the super-linear time over the size of the dataset (N) being analyzed, as it is needed by machine learning techniques [Blum1997], our FDR algorithm is linear on N and quadratic on the embedding dimensionality of the dataset. For more detailed information, see the Traina et al. publication to *SBBD 2000*.

3 Notable Outreach and Inclusion Activities

- Based on Informedia DLI-2 work and research directions, updated and delivered "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 "The Technology Review, Emerging Technologies that will Change the World: Data Mining", by M. Mitchell Waldrop, which appeared in MIT's *Technology Review,* Annual Innovation Issue Vol. 104, No. 1, February 2001. This article discusses rising technology fields, in this case video mining, that the magazine predicts will have a profound impact on the economy and on how we live and work:

> "Another hot area … is 'video mining': using a combination of speech recognition, image understanding and natural-language processing techniques to open up the world's vast video archives to efficient computer searching. For instance, when Carnegie Mellon University's Informedia II system is given an archive of, say, CNN news clips, it produces a computer-searchable index by automatically dividing each clip into individual scenes accompanied by transcripts and headlines."

- 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

A new company (Vivisimo, Inc.) has spun out of the Informedia project and CMU's Computer Science Department, with investment from the Commonwealth of Pennsylvania and a grant from the NSF SBIR program. The company is commercializing "document clustering" technology that: 1) gets inserted in a flow of search results, and 2) groups the search results into spontaneous hierarchical folders that are annotated with concise, informative labels. The mission of the company is to change how search results are delivered on computer screens everywhere, by replacing the current long, tedious lists with well-sorted, hierarchical folders.

5 Journal and Conference Proceeding Publications

Christel, M., Maher, B., and Begun, A. "XSLT for Tailored Access to a Digital Video Library", *JCDL 2001* (Roanoke, VA, June 2001).

Christel, M.G., and Warmack, A.S., "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.

Georgiadis, C., Triantafillou, P., Faloutsos, C. "Fundamentals of Scheduling and Performance of Video Tape Libraries", *Multimedia Tools and Applications Journal (MTAP)* (to appear).

Jin, R., Hauptmann, A. "Automatic Title Generation for Spoken News", *Human Language Technologies Conference (HLT2001)*, San Diego, California, March 18-21, 2001.

Jin, R., Hauptmann, A. "Title Generation for Machine-Translated Documents", *International Joint Conference on Artificial Intelligence (IJCAI01)*, Seattle, WA, August 4-10, 2001.

Korn, F., Pagel, Bernd-Uwe, Faloutsos, C. "On the 'Dimensionality Curse' and the 'Self-Similarity Blessing", *IEEE TKDE (Transactions on Knowledge and Data Engineering)*, special issue (best of ICDE2000), **13**(1), January/February 2001.

Meila, M., and Shi, J. "A Random Walks View of Spectral Segmentation", *International Conference on AI and Statistics (AISTAT) 2001*, Key West, Florida, January 4-10, 2001.

Meila, M., and Shi, J. "Learning Segmentation by Random Walks", *Neural Information Processing Systems (NIPS) 2000*, Vancouver, British Columbia, November 27-29, 2000.

Palmer, C.R., Pesenti, J., Valdes-Perez, R.E., Christel, M.G., Hauptmann, A.G., Ng, D., and Wactlar, H.D. "Hierarchical Document Clustering of Digital Library Retrieval Results", demonstration at *JCDL 2001* (Roanoke, VA, June 2001).

Pan, Jia-Yu, and Faloutsos, C. "VideoGraph: A New Tool for Video Mining and Classification", *JCDL 2001* (Roanoke, VA, June 2001).

Proietti, G., Faloutsos, C. "Accurate Modeling of Region Data", *IEEE TKDE (Transactions on Knowledge and Data Engineering)* (to appear).

Proietti, G., Faloutsos, C. "Analysis of Range Queries and Self-Spatial Join Queries on Real Region Datasets Stored Using and R-Tree", *IEEE TKDE (Transactions on Knowledge and Data Engineering)*, **12**(5), Sept./Oct. 2000, pp 751-762.

Santos, R.F., Traina, A., Traina, C., Faloutsos, C. "Similarity Search Without Tears: The OMNI Family of All-purpose Access Methods", *ICDE 2001*, Heidelberg, Germany, April 2-6, 2001.

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, C., Traina, A., Wu, L., Faloutsos, C. "Fast Feature Selection Using Fractal Dimension", *XV Brazilian Symposium on Databases (SBBD)*, Paraiba, Brazil, October 2000.

Valdes-Perez, R.E., Pericliev, V., Pereira, F. "Concise, Intelligible, and Approximate Profiling of Multiple Classes", *International Journal of Human Computer Systems*, **53**(3), 2000, pp. 411-436.

Villarroel, A. "Infrastructure for Delivering Fully-Searchable Streaming Media to the Internet", Master's Thesis, Information Networking Institute, Carnegie Mellon University, December, 2000.

Wactlar, H., and Derthick, M. "Accessing News Video Libraries through Dynamic Information Extraction, Summarization, and Visualization of Video Meta –Data", demonstration at *JCDL 2001* (Roanoke, VA, June 2001).

Wu, L., Faloutsos, C., Sycara, K., Payne, T.R. "FALCON: Feedback Adaptive Loop for Content-based Retrieval", *Proceedings of the 26th International Conference on Very Large Databases*, Cairo, Egypt, September 10-14, 2000, pp. 297-306.

Yi, Byoung-Kee, Faloutsos, C., "Fast Time Sequence Indexing for Arbitrary Lp Norms", *VLDB 2000*, Cairo, Egypt, Sept. 10-14, 2000.

6 Presentations, Demonstrations, and Industry Visitors

Aug 2000	Valdes-Perez: Presentation on clustering of search results to technical leaders (e.g., CTO) and business devlopment at Altavista, Inc., California, August, 2000. Contact: Mark Wozniak, Director of Business Development.
Sept 2000	Wactlar/Stevens: Visit to Warhol museum to discuss research collaboration.
Sept 2000	Wactlar: Richard Li, Chairman, Pacific Century Cyberworks. Informedia presentation.
Sept 2000	Wactlar: "New Directions for Video Information", invited talk at United Daily News, Taipei, September 13, 2001.
Sept 2000	Wactlar: Visit to Academia Sinica, Hong Kong. Joint research discussions with Chinese University of Hong Kong, Taipei, ROC and Hong Kong, SARC.
Sept 2000	Wactlar: ECHO Project European Commission Review Meeting, Luxembourg.
Sept 2000	Wactlar: "Informedia: Experience-On-Demand", Talk at DARPA EOD CVIM PI Meeting, Virginia Beach, Virginia.
Sept 2000	Christel: "Improving Access to Digital Video Archives through Informedia Technology", Christel, M., Audio Engineering Society 109th Convention Workshop on Digital Libraries, Preservation, and Metadata, Los Angeles, CA, September 23, 2000.
Sept 2000	Faloutsos: "FALCON: Feedback Adaptive Loop for Content-based Retrieval", VLDB 2000, Cairo, Egypt, September 10 - 14, 2000
Sept 2000	Faloutsos: "Multimedia Information Retrieval", Keynote talk, COMLEX 2000, Univ. of Patras, Greece, September 2000.

Oct 2000 Wactlar/Hauptmann: Participation in Talk Bank Meeting.

Oct 2000 Wactlar: "Approaching the Digital Human Memory Machine", Invited talk at NASA Design for Safety Conference, NASA Ames Research Center, Moffett Field, California.

Oct 2000 Wactlar: "Automated Indexing - New Theory and New Tools in Speech Recognition", Invited talk at FIAT/IFTA World Conference, Vienna, Austria.

Oct 2000 Faloutsos: "Massive Dimensionality: Issues and Solutions", Plenary talk, NSF-IDM workshop, Roanoke, VA, October 2000.

Oct 2000 Hauptmann: "Multimedia Information Retrieval from a Digital Video Library", Invited talk, Sixth International Workshop on Multimedia Information Systems, Chicago, Illinois, October 26-28, 2000.

Oct 2000 Shi: "Image Segmentation, Contours and Texture", Department Seminar, EPFL, Switzerland, October, 2000.

Nov 2000 Wactlar/Christel/Hauptmann: Szu-chia Lo, National Cheng-Kung University, Taiwan. Research discussion.

Nov 2000 Wactlar: Hong Kong SARC, Chinese University of Hong Kong and CMU Joint Project Kick Off Meeting for VIEW Consortium.

Dec 2000 Faloutsos: "Searching, Data Mining and Visualization in Multimedia Databases", Invited workshop speaker, DaimlerChrysler, Ulm, Germany, December 2000.

Dec 2000 Hauptmann: "Multimedia Video Mining", Invited workshop speaker, DaimlerChrysler, Ulm, Germany, December 2000.

Dec 2000 Shi: "Learning Segmentation with Random Walk", Neural Information Processing Systems(NIPS), Denver, Colorado, December 2000.

Dec 2000 Shi: "Image Segmentation: cuts, random walks, and phase space embedding", Computer Science Seminar, University of Chicago, Chicago, December 2000.

Dec 2000: Wactlar: "Capturing the Record of Human Experience in Video", Invited talk at APEC 2000, Asia-Pacific Economic Cooperation Workshop on Technology with Digital Museum (http://dm.ee.ntu.edu.tw/apec2000/default.htm) Academic Sinica, Taipei, R.O.C. December 8-9, 2000.

Jan 2001 Shi: "A Random Walks View of Segmentation", International Conference on AI and Statistics(AISTAT), Key West, Florida, January 2001.

Jan 2001 Wactlar: John Comack, Administrator, Kane Hospitals, joint research discussions.

Jan 2001 Wactlar/Hauptmann: Ellen Hughes and Bob Mitchell, Northrup Grumman. Research collaboration discussions.

Feb 2001 Wactlar: European Union Digital Library All Projects Meeting, Luxembourg (spoke on behalf of NSF).

Feb 2001 Wactlar/Hauptmann: Leland S. Kollmorgen, LCDR Dylan Schmorrow, DARPA

Feb 2001 Christel: Steven R. Vedro, Faculty Staff at University of Wisconsin, and Consultant for SRVedro Consulting, Telecommunications-based Strategies for Education, Economic and Community Development. Research discussion.

7 Other Report References

[Aberdeen1995] Aberdeen, J., Burger, J., Day, D., Hirschman, L., Robinson, P. and Vilain, M. "MITRE: Description of the Alembic System Used for MUC-6," *Proc. Sixth Message Understanding Conference (MUC-6)*, Morgan Kaufmann Publishers, Inc., Columbia, Maryland, 1999, pp. 141-155.

[Becker1987] Becker, R.A. and Cleveland, W.S. "Brushing Scatterplots," *Technometrics*, **29**(2), 1987. pp. 127-142.

[Berchtold1998] Berchtold, S., Böhm, C., Kriegel, H.-P., "The Pyramid-Tree: Breaking the Curse of Dimensionality," *Proc. ACM SIGMOD Conf. on Management of Data*, Seattle, Washington, 1998.

[Blum1997] Blum, A. L., and Langley, P., "Selection of Relevant Features and Examples in Machine Learning," *Artificial Intelligence*, **97**, 1997, pp. 245-271.

[Christel2000] Christel, M.G., Olligschlaeger, A.M, and Huang, C. "Interactive Maps for a Digital Video Library," *IEEE MultiMedia* **7**(1), 2000, pp. 60-67.

[Fayyad1998] Fayyad, U., "Mining Databases: Towards Algorithms for Knowledge Discovery," *Bulletin of the IEEE Technical committee on Data Engineering*, **21**, 1998, pp. 39-48.

[Gargi2000] Gargi, U., Kasturi, R., and Strayer, S.H., "Performance Characterization of Video-Shot-Change Detection Methods," *IEEE Trans. On Circuits and Systems for Video Technology*, **10**(1), February 2000, pp. 1-13.

[Houghton99] Houghton, R. "Named Faces: Putting Names to Faces", *IEEE Intelligent Systems Magazine*, **14**(5), September/October, 1999, pp. 45-50.

[Kubala1998] Kubala, F., Schwartz, R., Stone, R., and Weischedel, R., "Named Entity Extraction from Speech", *Proc. DARPA Workshop on Broadcast News Understanding Systems*, Lansdowne, VA, February 1998.

[Li2000] Li, F.C., Gupta, A., Sanocki, E., He, L. and Rui, Y. "Browsing Digital Video," *Proc. CHI '00*, Amsterdam, April 2000, pp. 169-176.

[Ng2000] Ng, T.D., "A Concept Space Approach To Semantic Exchange," Ph.D. Dissertation, The University of Arizona, April 2000.

[Pagel2000] Pagel, B.-U., Korn, F., Faloutsos, C., "Deflating the Dimensionality Curse Using Multiple Fractal Dimensions," *Proc. 16th Conf. on Data Engineering (ICDE)*, San Diego, CA, 2000.

[Roth97] Roth, S.F., Chuah, M.C., Kerpedjiev, S., Kolojejchick, J.A., and Lucas, P., "Towards an Information Visualization Workspace: Combining Multiple Means of Expression," *Human-Computer Interaction Journal*, 1997, **12**(1-2): pp. 131-185.

[Rowley1998] H. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(1), January 1998, pp. 23-38.

[Satoh1997] Satoh, S., and Kanade, T. "NAME-IT: Association of Face and Name in Video," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR97)*, San Juan, Puerto Rico, June 1997.

[Schneiderman2000a] H. Schneiderman, "A Statistical Approach to 3D Object Detection Applied to Faces and Cars," Ph.D. Dissertation, CMU Tech. Report 00-06, Robotics Institute, Carnegie Mellon University, May, 2000.

[Schneiderman2000b] H. Schneiderman and T. Kanade, "A Statistical Model for 3D Object Detection Applied to Faces and Cars," *IEEE Conference on Computer Vision and Pattern Recognition*, June 2000.

[Yoshihiko1999] Yoshihiko, G., Renals, S. and Williams, G., "Named Entity Tagged Language Models," *Proc. ICASSP 1999*, Phoenix, Arizona, pp. 513-516.

[Zhong2000] Zhong, Y., "Apply Multimodal Search and Relevance Feedback In a Digital Video Library," Master's Thesis, Information Networking Institute, Carnegie Mellon University, May 2000.

8 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: _____