Visualization for Large Collections of Multimedia Information

Dave Himmel
Graphics and Multimedia
Applied Research and Technology
Boeing Shared Services Group
P.O. Box 3707, MS 7L-43
Seattle. WA 98124
david.p.himmel@boeing.com

Mark Greaves, Anne Kao, Steve Poteet
Natural Language Processing
Applied Research and Technology
Boeing Shared Services Group
P.O. Box 3707, MS 7L-43
Seattle, WA 98124
mark.t.greaves@boeing.com
anne.kao@boeing.com
stephen.r.poteet@boeing.com

Abstract

Organizations that make use of large amounts of multimedia material (especially images and video) require easy access to such information. Recent developments in computer hardware and algorithm design have made possible content indexing of digital video information and efficient display of 3D data representations. This paper describes collaborative work between Boeing Applied Research & Technology (AR&T), Carnegie Mellon University (CMU), and the Battelle Pacific Northwest National Laboratories (PNNL). to integrate media indexing with computer visualization to achieve effective content-based access to video information. Text metadata, representing video content, was extracted from the CMU Informedia system, processed by AR&T's text analysis software, and presented to users via PNNL's Starlight 3D visualization system. This approach shows how to make multimedia information accessible to a text-based visualization system, facilitating a global view of large collections of such data. We evaluated our approach by making several experimental queries against a library of eight hours of video segmented into several hundred "video paragraphs." We conclude that search performance of Informedia was enhanced in terms of ease of exploration by the integration with Starlight.

Introduction

Boeing uses very large collections of data in the process of engineering and manufacturing commercial jet airplanes and in delivering complex military and space systems. The company also creates large amounts of information in the form of manuals and other documents to be delivered with these products. The need to control the cost of creating and accessing this information is motivation for new methods of indexing and searching computerized digital data.

In 1996, the Natural Language Processing Group at Boeing AR&T began a collaboration with PNNL to jointly develop the Starlight information visualization system. AR&T developed the Text Processing Toolset (TPT) to perform indexing and querying operations in high-dimensional document spaces, and PNNL developed software for 3D graphics presentations. Although Starlight has a rich visual presentation, it processes only text and static imagery, and lacks any inherent capability for extracting content from multimedia data. Concurrently, as part of ongoing research into digital libraries, the AR&T Multimedia Group began support of the CMU Informedia project and obtained a demonstration system for searching indexed digital video information. At the heart of Informedia is a subsystem that creates text metadata that is descriptive of digital video content. This suggested a way to integrate the two systems.

1 Informedia

The Informedia Project has established a large on-line digital video library, incorporating video assets from WQED/Pittsburgh. The project is creating intelligent, automatic mechanisms for populating the library and allowing for its full-content and knowledge-based search and segment retrieval. Our approach applies several techniques for content-based searching and video-sequence retrieval. Content is conveyed in both the narrative (speech and language) and the image. Only by the collaborative interaction of image, speech, and natural-language understanding technology can we successfully populate, segment, index, and search diverse video collections with satisfactory recall and precision.

The Informedia Project uses the Sphinx-II speech recognition system to transcribe narratives and dialogues automatically. The resulting transcript is then processed with methods of natural language understanding to extract subjective descriptions and mark potential segment boundaries where significant semantic changes occur. Comparative difference measures are used in processing the video to mark potential segment boundaries. Images with small histogram disparity are considered to be relatively equivalent. By detecting significant changes in the weighted histogram of each successive frame, a sequence of images can be grouped into a segment. This simple and robust method for segmentation is fast and can detect 90% of the scene changes in video.

Segment breaks produced by image processing are examined along with the boundaries identified by the natural language processing of the transcript, and an improved set of segment boundaries are heuristically derived to partition the video library into sets of segments, or "video paragraphs."

Figure 1 illustrates the video searching facilities of Informedia, which include:

- Filmstrip (lower-left of Figure 1) – Select thumbnail images to view video paragraph.
- Selective play (upper right of Figure 1) – Prev/next paragraph, prev/next term hit.
- Cursor browse – Abstracts and search terms available in filmstrip and play window.
- Text query (upper left of Figure 1) – Terms parsed from natural language query.
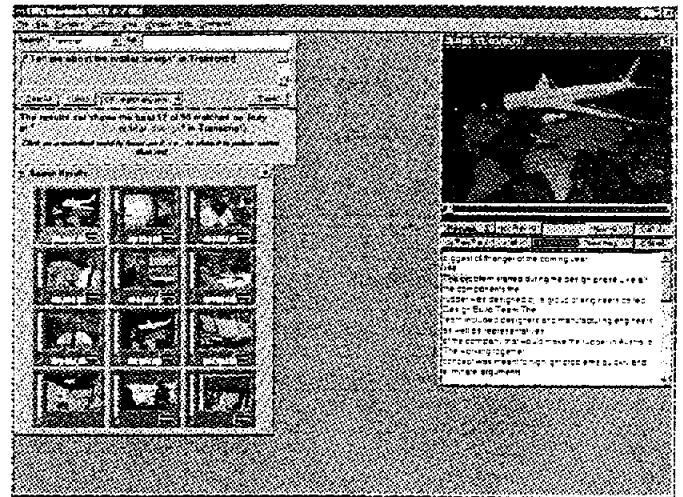- Skim (not shown) – View video in 10% of normal time.

Figure 1 – Informedia Search Screen

The reader can find more in-depth discussions of the Informedia project and technologies in References 1-4.

2   Starlight

Starlight was originally developed as an interactive information visualization environment for the US Army Intelligence and Security Command (INSCOM). It is designed to integrate several types of data (unstructured and structured text documents, geographic information, and digital imagery) into a single analysis space for rapid comparison of content and interrelationships (see reference 5). In this section, we will concentrate on the Starlight text processing and indexing functions.

A major problem with incorporating free-text documents into a visualization environment is that each document must be coded so that it can be clustered with other documents. The Boeing TPT is a prototype software engine that supports automatic coding and categorization of documents, concept-based querying, and visualization over large text document databases. The TPT combines techniques from statistics, linear algebra, and computational linguistics in order to take account of the total context in which words occur in a given document or query; it statistically compares a document context with similar contexts from other documents in the database. Through this technique, document sets can be represented in a way that supports visualization and analysis by the presentation components of Starlight.

The TPT performs two functions: First, it provides a powerful and flexible mechanism for concept-based searching over large text databases; second, it automatically assigns individual text units to coordinates in a user-configurable 3D semantic space. Both of these functions derive from the TPT core technique of representing large numbers of text units as points in a higher dimensional space and performing similarity calculations in this space. Conceptually, the flow of data through the TPT is diagrammed in Figure 2:
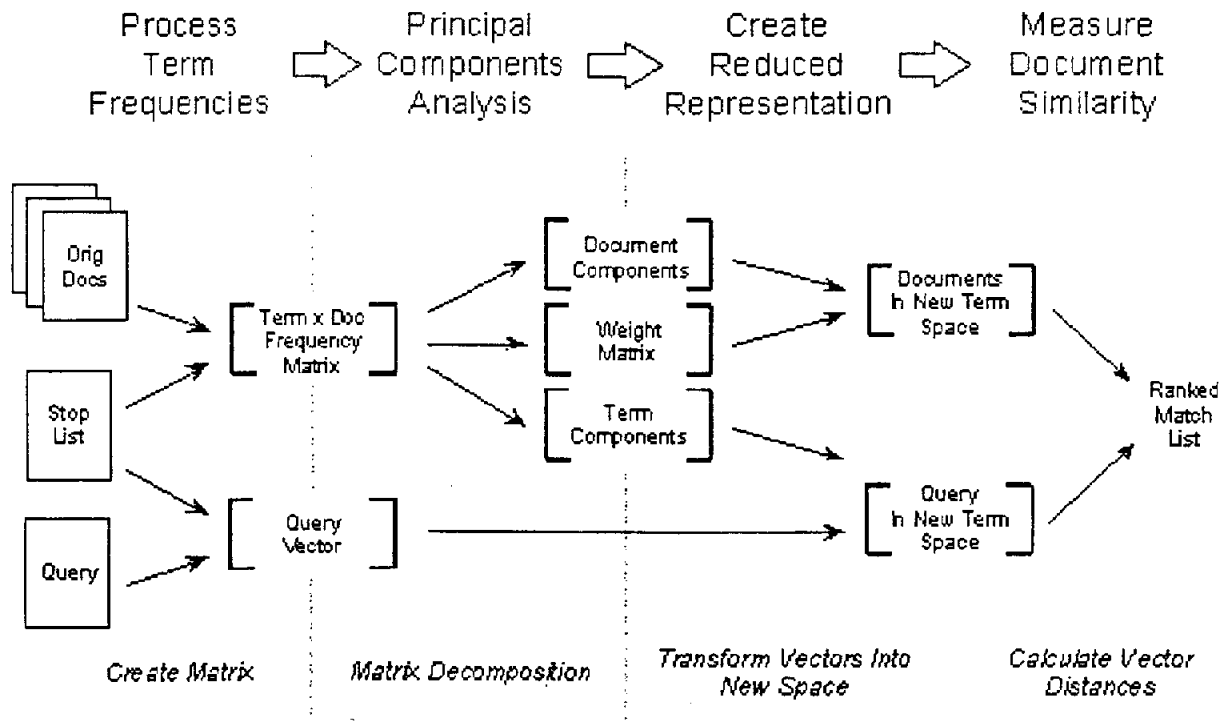

Figure 2 – TPT Processing Flow Diagram

Prior to TPT indexing, a text collection must be preprocessed in three stages involving manual intervention: First, the text is divided into units with topical granularity that best correlates with the expected query patterns. The units can be titles, subject lines, abstracts, individual paragraphs or an entire document. It can also be a caption or a piece of transcribed text from a video. Next, the text is "tokenized" into individual words and phrases. Finally, a list of "stopwords," or ignored terms, is chosen. These include determiners (*e.g.,* a, the), conjunctions (*e.g.,* and, or), and relatives (*e.g.,* what, which), and certain domain-specific terms.

After preprocessing, operation of the TPT indexing system is automatic (see Figure 2). The software builds a document/term matrix, performs several transformation and dimension reduction calculations, and stores output matrices in an object-oriented database for use by the Starlight visualization component.

Users of Starlight can visually explore the topical structure of a large text database by navigating through a 3D topic space where each item is represented as a point in a scatterplot. Items with close visual proximity have similar content. The TPT provides a selection of dimensions (with associated topic words), any three of which can potentially be selected for axes of the scatterplot.

Figure 3 shows a Starlight visualization screen containing a 3D scatterplot of the 322-item video metadata extracted from Informedia: each axis is labeled with the dominant topics measured by the TPT. At the right of Figure 3 is an example query with results.
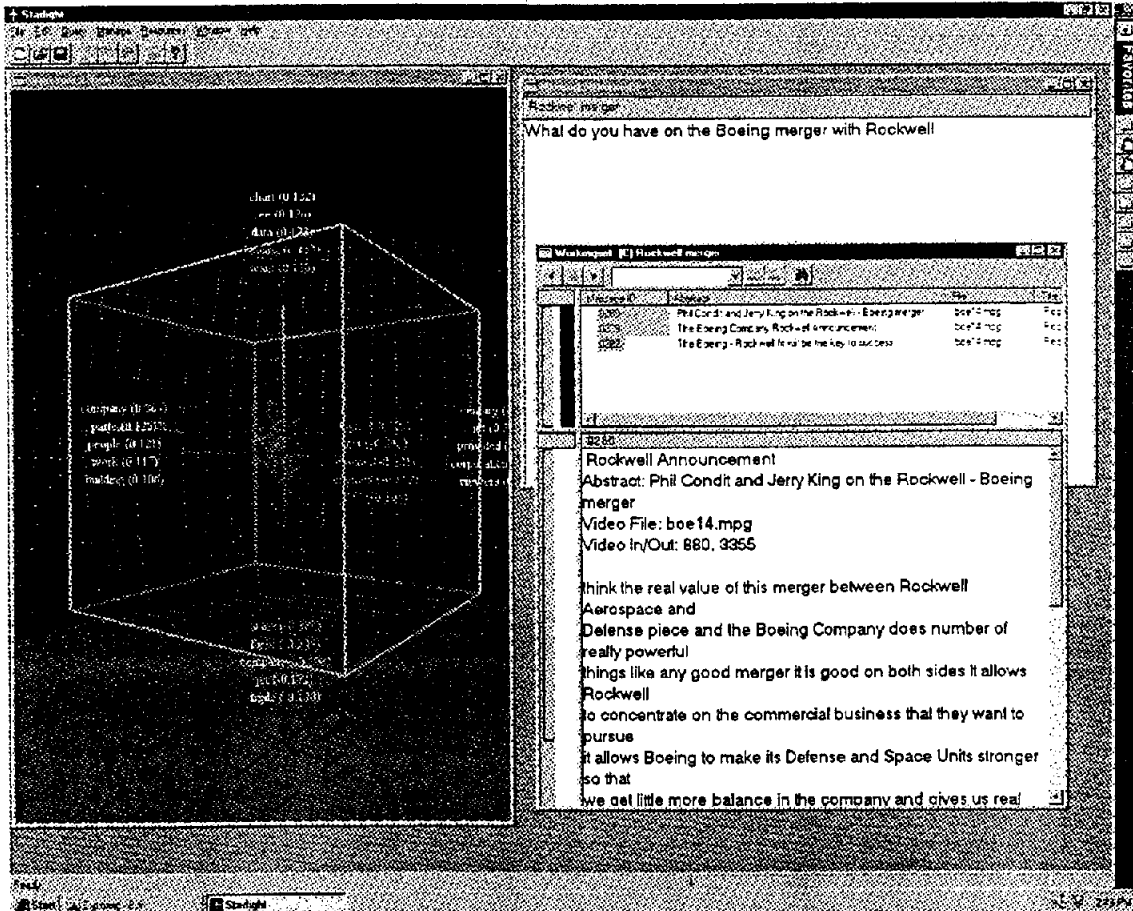

Figure 3 - Starlight Visualization Screen

3 Integration

Integration of the two systems required the insertion of two processing steps, each involving new software development. Figure 4 shows the flow of data in and between the two systems; new integration elements are shown as shaded boxes. The two key elements are: (1) Extract video paragraph text metadata from Informedia, and (2) Display selected video paragraphs.
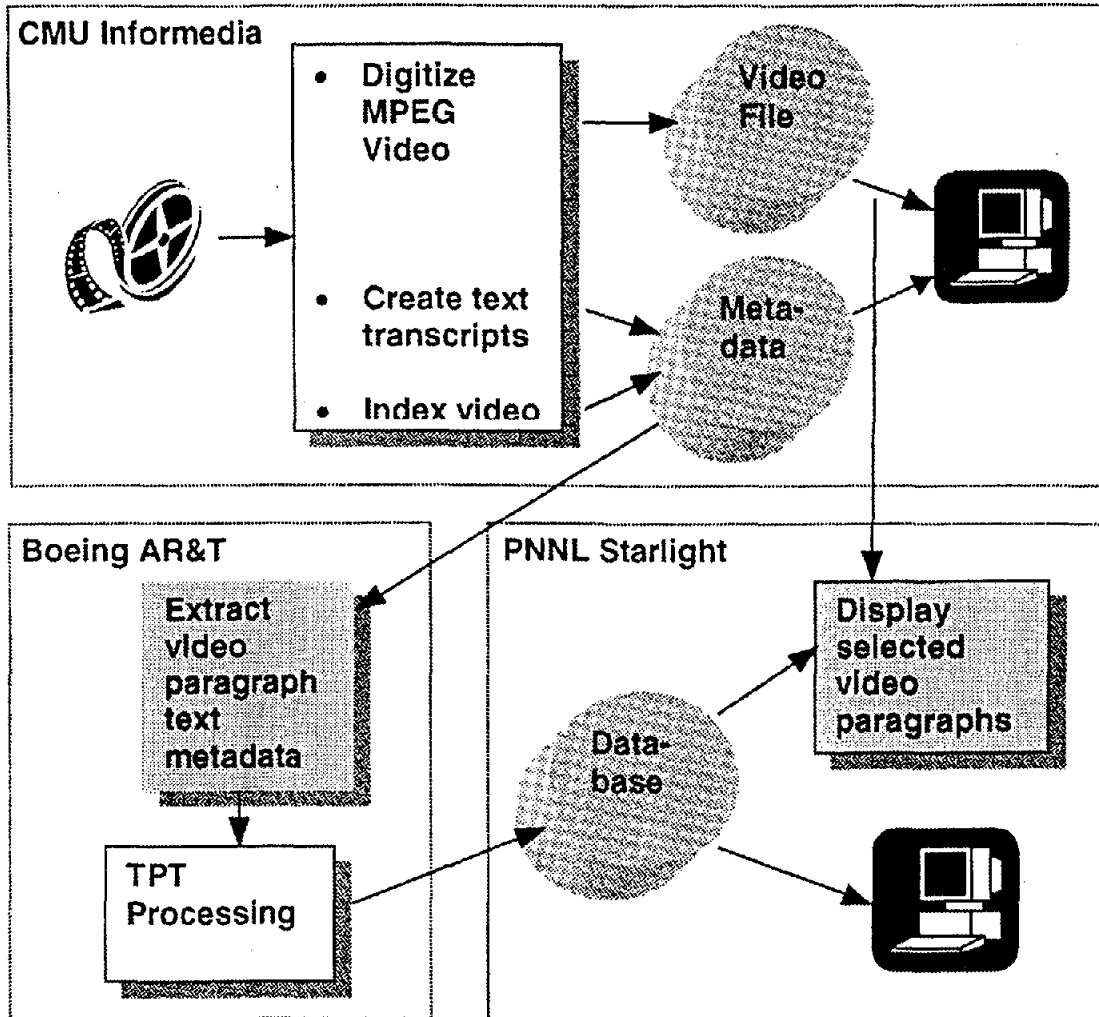

Figure 4 – Integration of Informedia and Starlight

Extraction of metadata is accomplished by a C program that reads ASCII text and control parameters from several files in the Informedia system and writes a collection of items compatible with TPT processing. Each item represents one video paragraph and includes the following fields:

- Video file name
- "In" and "Out" frames in the video file (for viewing the video paragraph)
- Title of video paragraph
- Abstract of video paragraph
- Transcript text

Display of video paragraphs occurs in the context of a web page containing a video viewer and the text transcript: Figure 5 shows an example. When the user selects a particular document within Starlight, a browser displays the HTML page. The browser was coded in Java and the MPEG viewer is a Microsoft ActiveMovie control. Each digital video file was kept intact. that is. the video paragraphs were not partitioned into separate files; rather, paragraphs are viewed by playing from the specified "In" frame to the "Out" frame of the appropriate MPEG file.
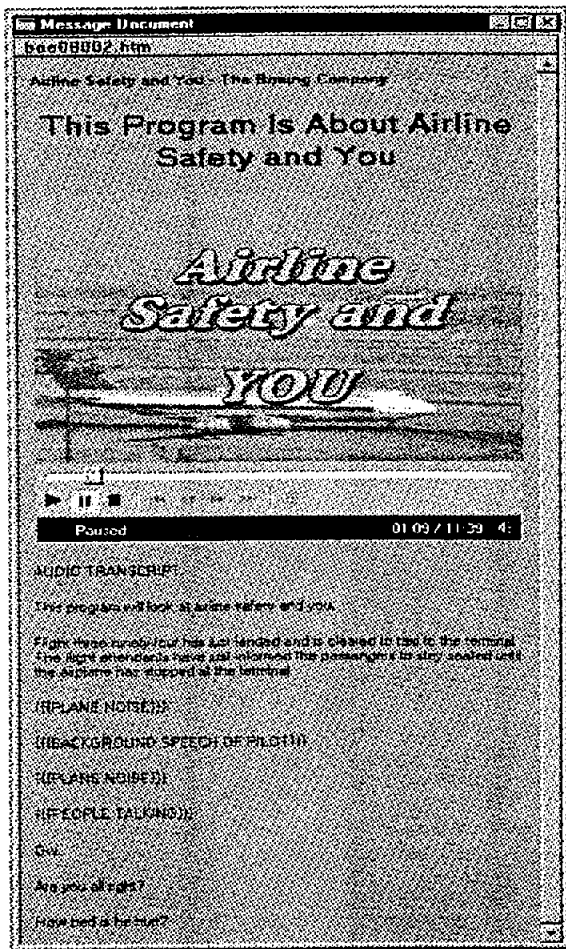

Figure 5 – Video Paragraph Display

4   Experimental Results

Although both Informedia and Starlight separately exhibit powerful capabilities for accessing large collections of data, it was interesting to speculate about synergistic

advantages. Neither system was designed for narrow, precise data querying. rather Starlight features all-inclusive views and Informedia facilitates iterative. progressive narrowing of focus. Motivation for integrating the systems was twofold: Give Starlight access to multimedia data, and investigate possible advantages of the global overview that Starlight can bring to the Informedia database.

We collected the following list of video titles from diverse sources at Boeing:

- PBS Feature: 21st Century Jet (Parts 1-6)
- Boeing Education Network (BEN) Tutorial: Excel Charting
- Boeing Education Network (BEN) Lecture: Computer Graphics
- Business Realities with Phil Condit: Cash Flow
- BTV News: Boeing Announces the Purchase of Rockwell
- Computer-Based Training: 757 Maintenance
- Maintenance Training: Airline Safety and You
- Maintenance Training: 777 Doors and Slides/Rafts
- Everett Factory Video: 767 Inboard Leading Edge

Altogether, this material comprises about eight hours of viewing. The digitized MPEG files occupy approximately 4.5GB of disk space and the metadata (transcripts, thumbnail images, and cross-reference files) is about 25MB in size.

The video data was indexed by Informedia, then we passed the resulting metadata to Starlight as described above. Starlight presents the entire collection of hundreds of video paragraphs as a scatterplot of points for viewing. which can be color-coded by video title. We immediately observed that items with like colors tended to cluster together, verifying that TPT processing was effective in bringing together topically related items.

In an experiment to evaluate the further effects of integration, we posed several queries to both Informedia and Starlight. noted commonalties and differences in response. then used the global

view of Starlight to find other items with similar content and to discover additional search terms. After Starlight reported query "hits," we observed the region in 3D space encompassed by the video file containing the most hits. Using the cursor brush to see abstracts, we examined all items (of any color) within this region, and noted interesting terms. Because the axes were topically labeled, additional terms were also available from the axis in the vicinity of the cluster. Table 1 shows the results of this experiment.

Table 1 – Experimental Results

Initial Query	Informedia Hits	Starlight Hits		Other Items in Starlight Region		Additional Search Topics Discovered
		Total	Same as Informedia	Total	Relevant	
How do I use the emergency exits?	6	5	4	10	8	Hazard, Safety, Test
What is fly by wire?	4	4	4	30	0	Plane, ETOPS, Flight deck, First flight
Tell me about scientific visualization.	12	8	7	26	12	Chart, See, Data, Misfit
What do you have on the Boeing merger with Rockwell?	4	3	3	12	3	Company, Parts, People, Working Together, Japanese suppliers, Austria suppliers

The last three columns indicate a positive result by showing additional video paragraphs and search terms relating to the initial query. The additional items varied in degree of relevance, but the cost of the new information is low in that only a few seconds were required to examine each. This illustrates the capacity of the integrated approach as an interactive tool for ready access to video content.

Conclusion

This effort clearly succeeded in the goal of providing access to multimedia content for the Starlight system. We have also shown the usefullness of Starlight global visualization of text metadata for video content.

The concept of bringing text metadata into Starlight is extensible to image, sound, animation, and other media, which suggests further experimentation with other forms of multimedia information and methods of generating metadata.

Both Informedia indexing and Starlight processing require some manual intervention. In order for these approaches to be efficient and cost-effective, we must develop fully automatic methods for creating and processing text metadata for multimedia information. It may be possible to do this by compromising the quality of metadata (perhaps by using unedited speech recognition from Informedia); a future experiment would be to attempt this compromise and discover the effect on search performance.

Acknowledgments

References

1. Wactlar, Howard D., Kanade, Takeo, Smith, Michael A., and Stevens, Scott M. *Intelligent Access to Digital Video: Informedia Project*, IEEE Computer, May 1996 pp46-52

2. Ravishankar, Mosur K. *Some Results on Search Complexity vs Accuracy*, Proceedings of DARPA Spoken Systems Technology Workshop, Feb. 1997

3. Christel, Michael G., Winkler, David B., and Taylor, Roy C. *Multimedia Abstractions for a Digital Video Library*, Proceedings of ACM Digital Libraries '97, Philadelphia, PA, July 1997.

4. Kanade, Takeo *Immersion into Visual Media: New Applications of Image Understanding*, IEEE Expert, February 1996, pp73-80.

5. Risch, John, May, Richard, Dowson, Scott, and Thomas, James *A Virtual Environment for Multimedia Intelligence Data Analysis*, IEEE Computer Graphics and Applications, November 1997.