

Exploring Cultural Origins and Evolution by Navigating Searchable Video and Film Archives

Howard D. Wactlar
Carnegie Mellon University, Pittsburgh, Pennsylvania, U.S.A.
wactlar@cmu.edu

Introduction

Vast collections of film, video and audio recordings which have recorded events of the last century and portrayed all of history since time began, remain a largely untapped resource of historical, cultural and scientific value. The Informedia Digital Video Library project at Carnegie Mellon University has pioneered new approaches for automated video and audio data storage, indexing, search and retrieval and embedded them in a system for use in education, training, information and entertainment [Wactlar00,99,96]. This project was initiated in 1994 as one of six national Digital Library Initiative (DLI) projects in the United States. The newest phase of this effort introduces new paradigms for video information access and understanding, and provides metrics to evaluate them. We aggregate and integrate video content on-demand to enable summarization and visualization that provides the users with coherent responses to their queries in a useful broader context, perhaps with historic or geographic perspectives.

Background

The Informedia system provides “full-content” search and retrieval of current and past TV and radio news and documentary broadcasts. The system implements a fully automated process to enable daily content capture, information extraction and storage in on-line archives by applying artificial intelligence and advanced systems technology. The current library consists of approximately 3,000 hours, in a two terabyte library of daily news captured over the last four years and documentaries produced for public television and government agencies. This prototype database allows for rapid retrieval of individual “video paragraphs” which satisfy an arbitrary spoken or typed subject area query based on the words in the soundtrack, closed-captioning or text overlaid on the screen. There is

also a capability for matching of similar faces and images.

Our approach uniquely combines speech recognition, image understanding, and natural language processing technology to automatically transcribe, segment and index the linear video. These same tools are applied to accomplish intelligent video search, navigation and selective retrieval. The process automatically generates various summaries for each story segment:



Figure 1. IDVL interface showing 12 documents returned for “El Niño” query along with different multimedia abstractions for certain documents.

headlines, “filmstrip” story-boards and “video-skims”.

Informedia used automatically computed descriptors of the video library, i.e., metadata, to return a set of video documents in response to a user query, as shown in Figure 1. Further responsibility for interpreting those results was left

to the user. Informedia-II, the second phase of this research, enables users to more quickly interpret and assimilate information relevant to their needs via automatic, intelligent synthesis of distributed video sources. Summaries rather than documents become the units of discourse, as shown in Figure 2. Video sources can be viewed in the context of these summaries, showing how events unfold over time and across geographic boundaries.

Spatial and Temporal Analysis

Keywords were used successfully for information retrieval within Informedia, but were less effective for cases such as distinguishing the person named “Prince William” from the location of “Prince William Sound.” The Informedia-II Project automatically extracts references to named entities from the video material, i.e., names, places, dates and times. The interface then builds summaries from the correlations and aggregations of such named entities. For example, a video report about Prince William Sound can be automatically associated with other Alaskan waterways through geographical thesauri. A user searching for oil spills could be shown a map interface with highlighted “hot spots” where each hot spot is a cluster of relevant documents in a particular geographic region; one such cluster would be on this Alaskan waterway.

The original Informedia could only sort results by recency or relevance. In Informedia-II, extracting temporal information provides the ability to analyze results relevant to a query based on the period mentioned or depicted within them, minimizing the effects of redundancy in the content and exposing trends and developments. Given a collection of stories covering an evolving event, the system will return a time-sequenced set of document segments derived from multiple sources with a visual timeline that summarizes the information. The sources may be text as well as audio, video or still images. Viewing a sequence of events with both geographic and time perspectives may enable the user to detect trends and infer causation across years or centuries.

An interesting example to consider might be the results of a query about “Chinese inventions” drawn from an international collection of historical

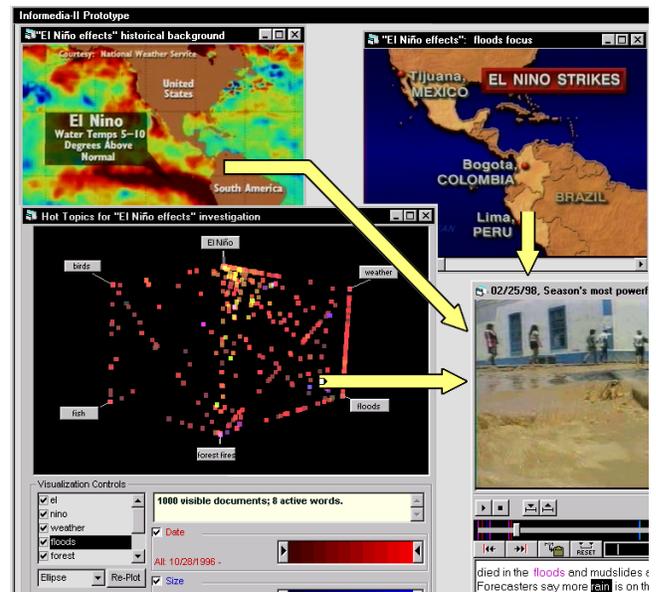


Figure 2. Additional views provided by the Informedia-II interface for an inquiry into “El Niño effects.”

documentaries. The time lag between invention in China and application in the West may be revealed automatically on the generated timelines and traced geographically on the generated maps. For example, gunpowder, invented around 850 AD, was not associated with tales of fighting in Europe until those of the 13th century, but was used for military purposes well before that by the Arabs, who may have become acquainted with it as a result of their active trade with China. The magnetic compass, invented in China around the 4th century BC, did not appear in stories of European adventure and trade before those depicting the early 12th century AD.

Summarization and Visualization

Informedia offered static text summaries of individual video paragraphs, static filmstrip summaries of individual video paragraphs, and precompiled static video “skims” of the content in a video paragraph [Christel00,98]. For Informedia-II we build text, image and video abstractions that are dynamically computed based on the content and the expressed preferences of the user.

Moreover, summaries extend across large numbers of video documents to represent events across time and space, illustrated in Figure 2. Instead of using a single paragraph as the unit of analysis, we will ultimately create synthetic video documents that

reflect the distillation of information across multiple video paragraphs.

Beyond the video medium, we will create an interactive screen that summarizes documents from text corpora, images, audio and video in one single abstraction, with the ability to drill down to any refinement in period or location and to any level of detail. Given multiple, near-duplicate or overlapping units of information, the system's goal is to combine them into one unit, a synthetic "story" or video magazine that summarizes all the salient information.

Various visualization techniques have been designed to help the user obtain an overview of the temporal and spatial relationships given large sets of documents in multiple media. Distribution of the video library's contents across time, space, topics and perspectives are exposed for efficient user examination and action. Details can be selectively shown in the context of these dimensions, i.e., the visualizations are active objects supporting direct manipulation for zooming into specific areas. Figure 2 shows summaries and visualizations for "El Niño effects." By zooming in on a contributing document for the summary, a video on Peruvian flooding is shown.

References

[Wactlar00] Wactlar, H., Olligschlaeger, A., Hauptmann, A., Christel, M., "Complementary Video and Audio Analysis for Broadcast News Archives", *Communications of the ACM*, 43(2), pp. 42-47, February, 2000

[Wactlar99] Wactlar, H., Hauptmann, A., Gong, Y., Christel, M., "Lessons Learned from the Creation and Deployment of a Terabyte Digital Video Library", *IEEE Computer* 32(2), pp. 66-73, 1999.

[Wactlar96] Wactlar, H., Stevens, S., Smith, M., Kanade, T., "Intelligent Access to Digital Video: The Informedia Project", *IEEE Computer*, Vol. 29, No.5. Digital Library Initiative special issue, May, 1996.

[Christel00] Christel, M., Olligschlaeger, A., Huang, C., "Interactive Maps for a Digital Video Library", *IEEE Multimedia* 7(1), pp. 60-67, 2000.

[Christel98] Christel, M., Martin, D., "Information Visualization Within a Digital Video Library", *Journal of Intelligent Information Systems* 11(3), pp. 235-257, 1998.