

Informedia-II: Search and Summarization in the Video Domain

INTRODUCTION

Vast collections of video and audio recordings which have captured events of the last century remain a largely untapped resource of historical and scientific value. The Informedia Digital Video Library project at Carnegie Mellon University has pioneered new approaches for automated video and audio indexing, navigation, visualization, search and retrieval and embedded them in a system for use in education, information and entertainment environments. This project was initiated in 1994 as one of six Digital Library Initiative (DLI) projects funded jointly by NSF, DARPA and NASA, and is the only one focused on the video medium. We continue the fundamental goal of enabling for video all the functionality and capability existing for textual information retrieval, while leveraging its temporal and visual qualities for richer information delivery. Informedia-II establishes an era focused for the user as we introduce new paradigms for video information access and understanding. We aggregate and integrate video content on-demand to enable summarization and visualization that provides responses to 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 a 1,500 hour, one terabyte library of daily news captured over the last two 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: headlines, *filmstrip* story-boards and *video-skims*. Figure 1 illustrates a typical query and result set display.

The Informedia-II Project improves both speed and accuracy of the underlying information extraction, now including interpretation of name, place, date and time references, and adding the challenges of dynamic story segmentation, speaker voice and face identification, and video event characterization and similarity matching. The performance goals include real-time processing for analysis to enable contemporaneous incorporation into an active library, and interoperability across distributed proprietary video archives.

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



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

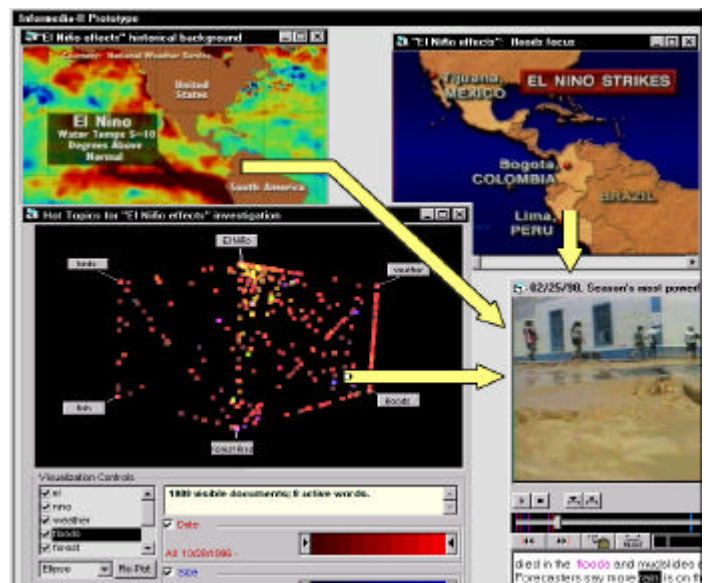


Figure 2. Additional views provided by the Informedia-II interface for an inquiry into "El Niño effects."

events unfold over time and across geographic boundaries, allowing visualizations that emphasize time and space perspectives.

SPATIAL AND TEMPORAL ANALYSIS

Keywords were used successfully for information retrieval within Informedia, but were less effective for cases such as distinguishing the person “Prince William” from the location “Prince William Sound.” The Informedia-II Project will automatically extract references to named entities from the video material, i.e., names, places and times. The interface will then build from the correlations and aggregations of such named entities. For example, a video report about Prince William Sound can be associated with 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.

Whereas Informedia merely sorts results by recency (or relevance), extracted temporal information in Informedia-II will provide the ability to analyze search results over time, minimizing redundancy 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 or video, and may be accessible from multiple, distributed content providers.

SUMMARIZATION

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. For Informedia-II we build text, image and video abstractions that are dynamically computed based on the content and history of the user queries and interactions. Moreover, summaries can extend across 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 create synthetic video documents that reflect the distillation of information across multiple video paragraphs.

Beyond the video medium, we propose to create *collages* that summarize documents from text corpora, images, audio and video in one single abstraction. Given multiple, near-duplicate or overlapping units of information, the proposed work will combine them into one unit, a *synthetic story* or video magazine that summarizes all the salient information.

MULTIMODAL EXPLORATION

We will allow multidimensional queries that may combine image elements, video clips, text and speech. New visualization techniques will be designed to help the user obtain an overview of the temporal and spatial evolution

given large sets of documents in multiple media. Distribution of the video library’s contents across time, space, topics and perspectives will be 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.

IMPACT

The Informedia-II Project has the potential to unlock the barriers to productive video information access. Via automated processing, the contents of traditional video libraries can be more fully exposed where before the only window to that data was through the titles and perhaps a few manually entered keywords. Videos of past reviews, training and teleconferences can become accessible corporate assets. Outreach for national archives and public service video collections can be enhanced.

We envision that the legacy for this project will enable:

- Broader accessibility to the historical video record being created across the globe.
- Pervasive societal impact as we provide a set of new capabilities to understand how events evolve and are correlated over time and geographically.
- Increased use of video in domains previously dominated by text or speech.

We will work with content providers to make their materials more accessible, and to study patterns of use of their video by appropriate communities of users. Our goal for unlocking the information embedded in video for easy access by the student, teacher, journalist, scientist or home user has the potential to create video resources with significant educational and commercial value. Our approach of automatically processing large libraries will stress current limits on network and i/o bandwidth, disk space and processor speed. Our focus on video as a searchable resource has broad implications for information gathering and dissemination.

AFFILIATES

The Informedia Project is actively seeking partners to help achieve the vision of simplifying information discovery in the video domain. An Informedia Affiliates program has been established for content producers, media distributors, and technology developers. Carnegie Mellon University is uniquely positioned to pursue this work, given its rich history of speech recognition, image processing and information retrieval work and the past successes of the Informedia Project. For further information, contact Howard D. Wactlar, wactlar@cmu.edu, 1+412-268-2571, Carnegie Mellon University, Pittsburgh, PA 15213.

