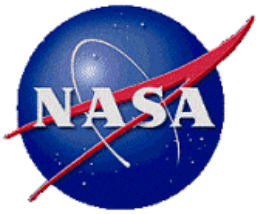

WHITE PAPER

Video Indexing and Streaming of NASA Multimedia Content

Developed at the
NASA Goddard Space Flight Center

by
Patrick Healey, Library Information Services, Code 292
& Ed Rogers, Systems Management Office, Code 300

May 17, 2004

Introduction

NASA has a wealth of human capital in its historical film and video collections that has, for the most part, been an underutilized resource. Because of the difficulty in accessing the knowledge locked within these audio and video files, they have not been widely considered a valuable information resource. However, advances in streaming digital technologies, specifically, internet bandwidth speeds, reliability and developments in digital video and audio encoding, and the ability to index segments of digital recordings have made it possible to open these resources to a vast number of users.

In response to this, the NASA Goddard Space Flight Center's (GSFC) Library and GSFC Knowledge Management Office have collaborated to produce the Center's "Streaming Media Center." Newly developed tools and applications are presently deployed, or in development, for the center to further expand the usefulness of these technologies, including indexing, metadata, video interval retrieval by keyword searching, and synchronized PowerPoint slides. All of these aid the users ability to sort through this abundant information source, and allow on-demand desktop delivery of previously unused content.

Streaming Media

Streaming Media refers to the synchronized delivery of audio and video content, on the Internet, which is received as an uninterrupted real-time stream to a user's desktop. To make this transition, the stream is encoded or "digitized" from a given source, be it a camera for a live event, video tape or in some cases a optical disk (CD or DVD) into a new format, that is suitable for streaming, through lower bandwidth connections indicative of the internet. Distribution, for the stream, is broadcast by a server application that acts as an interface for users requesting access to the streamed video. The advantage of this method of distribution assures that not only can multiple users access the stream since servers are constructed to handle a larger number of users, but servers can also be utilized as an distribution site, for previously recorded video as well. An additional feature of this configuration is the ability to link multiple servers, to avoid network overloading and also to allow an even greater number of simultaneous users.

Accessing a stream, from a server, requires that the user has previously installed an appropriate media player, capable of decoding the incoming streaming format. The media player receives the data as a series of packets of compressed data from the server, and then decompresses these packets back into a format that is readable by the player. Media players can start playing back or displaying video and audio as soon as enough data has been received and stored in the receiving station's buffer. Most media player software is available as a free download from the format developer. Windows Media Player, Real Player and QuickTime Player are the most common applications today. This medium makes it possible for teams to share expert presentations widely, review older digital fiels and historical records and to schedule more flexibly (files may be viewed on a delayed basis). Also conferences, key meetings, reviews and decision points can be videoed and subsequently reviewed easily from distributed points in the Enterprise.

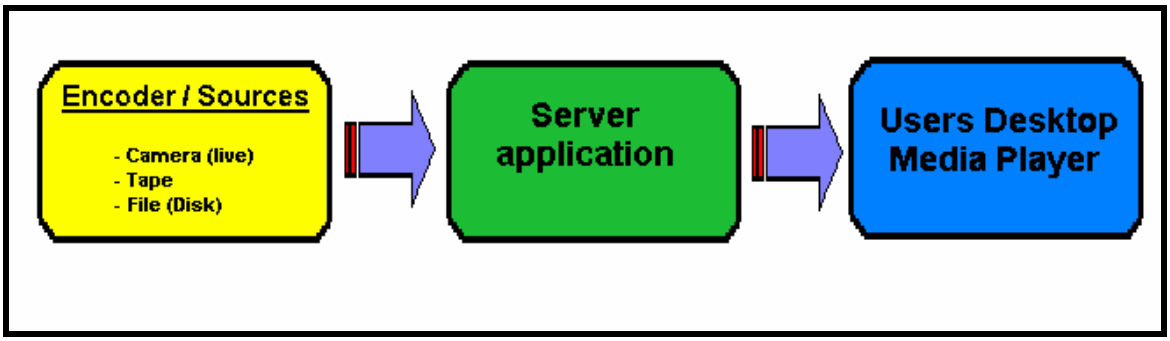

Fig. 1

Encoding (capturing)

Encoding is the processing of video from a given source to a format that is suitable for streaming. Common sources include live camera feeds, VHS Tape and in some instances, DVD and CD's. An encoder is simply a desktop computer that has a video capture card and encoding software installed and configured for the video capture card. The Goddard Library utilizes Windows Media 9 Encoder software and Osprey video card's for all video capturing and streaming. Our encoded files are recorded in two bitrates simultaneously, 256Kbps at a frame size of 320x240 for broadband connections, and 56K 160x120, respectively, to accommodate dial-up connections from off center. On center, there are two remote locations in which have encoders in place; the Building 8 and 3 auditoria, which are maintained by the Technical Information and Service Branch (TISB), Code 293. From these locations, we are able to record events live, there by reducing the turn around time needed to make the video available. In addition, events from these locations can also be streamed, with prior notification and approval. The encoding of all other content, from tape or disk, having utilized the same encoder setting previously mentioned, are processed from two computers located in the library. Encoding is the most time demanding aspect of the process due to the need to record in real-time: for example, one hour of video takes one hour to encode. An important clarification worth noting is that encoding of video, regardless of its source, is time intensive rather than labor intensive.

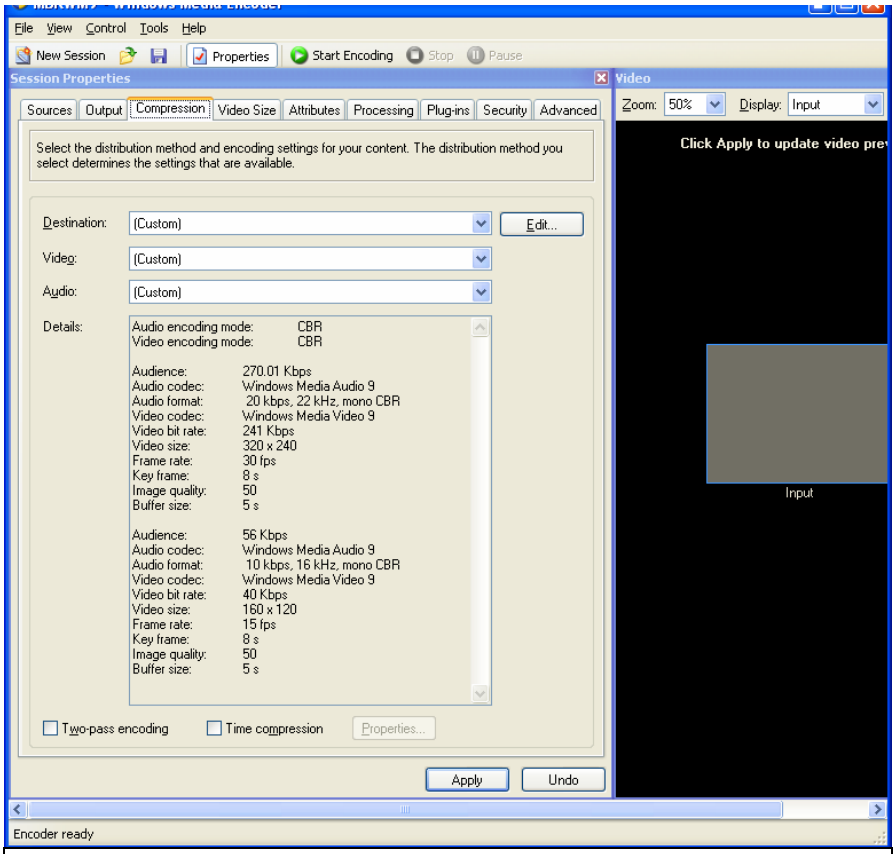

Fig. 2

Keyword Indexing Video Files for Search and Retrieval

There are two different methodologies for keyword indexing of online video: Speech-to-text conversions with subsequent indexing performed on the transcript, and speech recognition that utilizes a stored vocabulary for matching terms with the spoken words from the audio source. The GSFC Library has evaluated these primary means of indexing and has chosen the later of these two systems for its online video resources. There are inherent problems with the speech-to-text technologies currently available. The foremost is in the speech-to-text conversion software, which is the core of most streaming media indexing suites. Our inquiry into the accuracy of this conversion lead us to 3 conclusions.

1. Speech-to-text software conversion, and its subsequent transcript are not accurate due factors such as a speakers accent, broken English and even differences in harmonics between male and female voice ranges. What is misleading here is that there are a number of popular software products available on the market today that will allow a user to vocalize their intentions to their desktop computers, and that will transcribe their spoken words in to text. However, in order for this to work properly, the user must train the software to their particular voice and harmonics. Therefore, this is not plausible in our case, since our content is not from a single speaker, but rather a multitude of speakers form different genders, nationalities and dialects.

2. Indexing, based on poor transcriptions would result in inaccurate query response for users.

3. Due to the poor results, of the speech-to-text conversion process, considerable time would be required to edit the transcripts, therefore defeating our goal of automating the process as much as possible.

4. Most popular indexing suites, of this nature return only fixed intervals before and after a matched term. (For example: 10-20 seconds prior to, and after, a matched term)

The Audio Visual Search Engine

The GSFC Library has determined that the best application for our need is met through a search and retrieval suite (the Audio Video Search Engine, or AVSE) developed by a company called StreamSage. Our evaluation and ultimate adoption of this application was derived by a number of factors we determined to be crucial to the successful deployment of our video content:

1. The StreamSage software does not rely on speech-to-text conversion, but rather on the contextual analysis of speech recognition output, there by avoiding errors in the transcription process.

2. Due to the specific terminology used with NASA, the NASA Thesaurus was integrated into the core vocabulary of the AVSE in order to reflect the scientific

terminology commonly used in the Sciences and Engineering disciplines of the agency.

3. In addition to the expanded terminology acquired by the NASA Thesaurus, the AVSE understands the relationships between terms and their synonyms. This enables the AVSE to automatically expand query results and increase the likelihood of better relevant interval creation.

4. Query results are not restricted to fixed intervals. The AVSE uses speaker inflection (natural pauses in speech), language structure analysis and a large word model, to contextually index audio/video content. This results in the creation of contextually related interval lengths that allow the user to better understand the intended usage of a term spoken by the speaker.

5. Automation: Analysis and indexing of Goddard Library's specialized content requires little human interaction.

6. The ability to create custom profiles. Most of the required data entry is already included in the element fields, when inserting a video into the queue.

7. The StreamSage suite is adaptable to various software applications, including Accordent (PowerPoint integration with the corresponding video interval), as well as the Goddard Core Metadata Server being developed by the GSFC Library and the University of Maryland

Processing Video Content with the AVSE

The actual processing of content, into the StreamSage indexing database requires little effort by the library staff, beyond the preparation of the video. As seen in figure 3, the most labor-intensive aspect of the processing, lies in the video preparation, and not in the actual indexing procedures. The steps are as follows:

1. Content must first be encoded, edited and posted on the media server for streaming.

2. All video metadata and streaming location and protocols are added into the database.

3. Multiple videos can be "queued" in the administrator interface allowing for continual processing of content.

4. All processing through the indexing and speech recognition is automated.

5. The database records generated for each video are tagged using XML protocols.

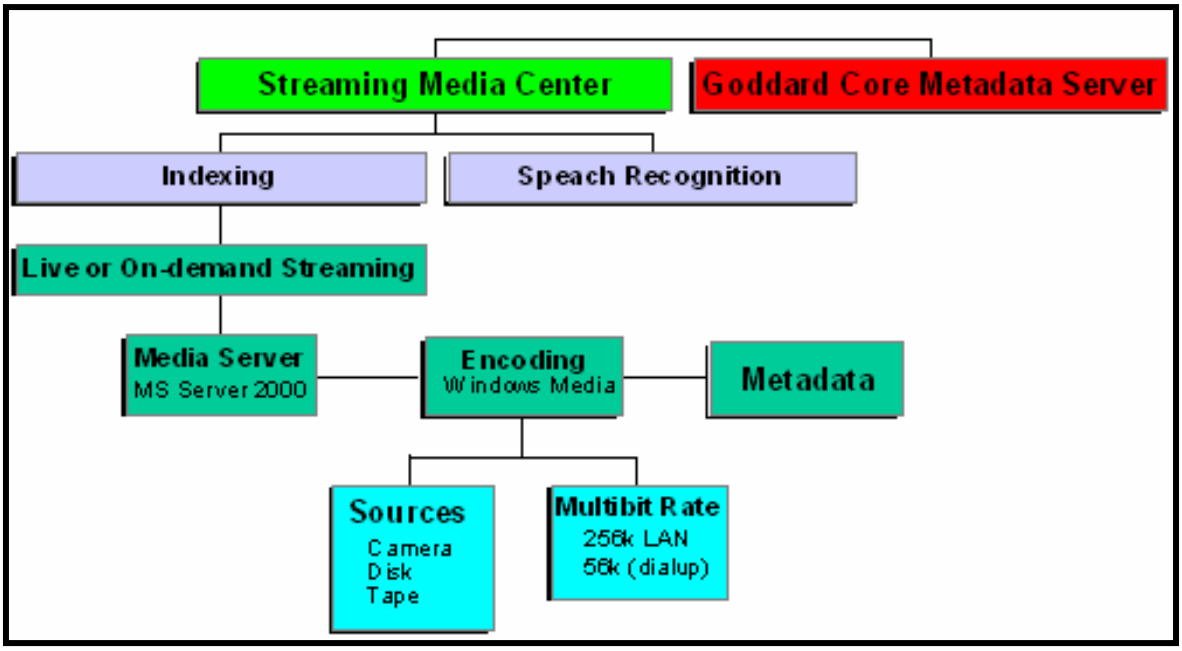

Fig.3

Applicability to NASA

The Goddard Library has started collecting video material for processing into the AVSE. Having hours of video available on line in a digital format that can be streamed to the desktop is valuable. Having that volume of video searchable by intelligent systems that index it into usable portions has enable people to actually access and use the valuable video content. This paper's position is that without the combination of these two support functions the vast majority of video resources will remain in obscurity. Since NASA has many video resources that might prove valuable from the past and will probably continue to generate many new video resources, it would be wise to consider setting up this capability and enabling its use across the Agency. The knowledge that video can and will be reused might encourage its more wide spread use in reviews, outside expert opinions and information distribution needs. The technology is now available to make video resources much more valuable than simply records of events.

Goddard is still working on proving value and utility from combining PowerPoint charts with video for presentation searching. This capability is currently unavailable in the Agency. As a result, much information contained in charts and presentation is not retrievable in any format. Video that can be indexed with embedded charts that are searchable may be the fastest way to include the many key presentations in the Knowledge Base for a project, program or enterprise.