A System for Customized News Delivery from Video Archives[1]

G. Ahanger and T.D.C. Little
Multimedia Communications Laboratory
Department of Electrical and Computer Engineering
Boston University, Boston, Massachusetts 02215, USA
(617) 353-9877, (617) 353-6440 fax
{*gulrukh,tdcl*}*@bu.edu*

MCL Technical Report No. 06-06-1997

Abstract– Video is a powerful medium for disseminating news as information. Like any other information, techniques are required to help search and locate interesting video content. In this paper we consider the construction of a system to automatically assemble an individualized news video when a sufficient volume of video is available for composition. We propose a vocabulary and representation for composition of news video. A model is presented that is used to store metadata and customize newscasts. The model facilitates representation of information clustering, object stratification, concept hierarchy, temporal ordering, and news "tours." The proposed customized news delivery system attempts to achieve a level of composition found in existing television news delivery.

Keywords: news delivery, video model, indexing, information composition, retrieval.

[1]In *Proc. 4th Intl. Conf. on Multimedia Computing and Systems*, Ottawa, Canada, June 1997. This work is supported in part by the National Science Foundation under Grant No. IRI-9502702.

1 Introduction

To reduce the onslaught of digital information, current research in information filtering and retrieval focuses on techniques for effective delivery of desired information. Each day new information is generated and a user is required to wade through this mire of content to access topics of interest. News information poses a major problem as news generation is a continuous process, originates from multiple sources, and due to it's dynamic nature results in a number of frequent revisions and versions of the same news. This problem can be alleviated to a certain extent by *information filtering*. Depending on a user's requirements the information is filtered at the source.

A number of systems have been developed [6, 9] for real-time presentation of text-based news. These systems utilize information filtering techniques on text-based news data such as USENET articles. Excite's NewsTracker [7] is one example of the many applications that provide text based news via the World Wide Web. In the video domain, analysis of closed-caption data leads to query on video data. The work by Brown et al. [3] is representative of this approach.

The time of presentation of the news plays an important role since news can age very fast. Hence, to provide a user with news in video format a news provider needs to expedite the editing process. A video should be annotated based on a model that will include features to provide fast access and automate customization of news according to a user's specifications. We need to utilize *information extraction* for this purpose. Information extraction systems emphasize the acquisition of facts from the incoming data (Fig. 1). Therefore, while servicing the needs of the community a news provider can cater to a single user with a specific need. The problems in adequacy of information, query representation and information organization should be addressed, rather than assuming (as in information filtering) that profiles of a user or community can provide effective and efficient techniques for video information composition [1].

We seek to automatically customize and deliver news video by combining segments from various sources while maintaining continuity. To achieve this goal we need to understand how different concepts in television news are combined to provide a user with the required news information. News items are composed of fragments of events, commentaries, and interviews which makes them good candidates for automatic customization.

In this paper we define an ontology which is used as a basis to extract information

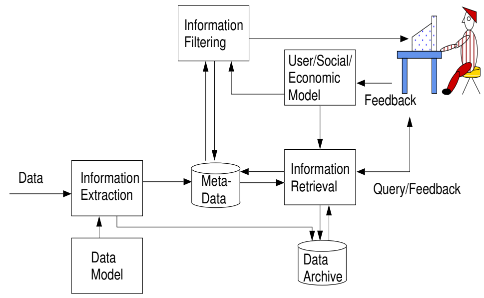

Figure 1: Block Diagram of the News Video Data Processing

from the news video data. We present a news video data model that can be used either to model a pre-composed newscast or to build a newscast from available information. In contrast to existing works involving text, we achieve customization by manipulating news video data based on existing techniques in broadcast news. Newscasts can be composed by the selection of logical segments from a video archive based on the ontology. We present a relational database schema which represents the semantic information that supports the relations between different media types defined in the data model. We also present techniques for using closed-caption data for customizing news video items.

The remainder of this paper is organized as follows. In Section 2 we define concepts which we use as fundamental building-blocks for automatic news item composition and demonstrate how we achieve this. We also present a relational database schema which defines the relationship between these concepts. In Section 3 we discuss the system architecture and the processes involved in achieving our objective. Section 4 concludes the paper.

2 A Newscast Video Data Model

Existing newscast data are pre-orchestrated, usually consisting of a headline followed by segments of news items. There is a logical flow of information in the news items and their position in a newscast depends on their relative importance or type of information they convey (e.g., politics, weather, sports) [12]. Each news item consists of an anchor person or reporter discussing the news occurrence and may also have field footage (recording of actual event occurrence) as shown in the Fig. 2. A newscaster acquires information from multiple sources and during presentation, and this wealth of information is integrated to provide a well structured semantic unit in a short duration. For example, different shots belonging to different instances of the same news item (different sources) can be composed together based on a story line. We require an information model that will satisfy these scenarios.

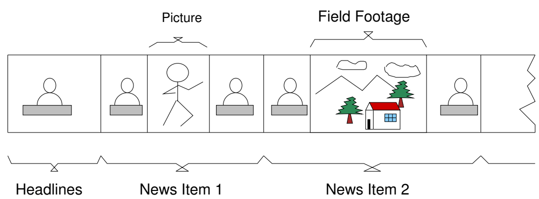

Figure 2: Content Arrangement in a Typical Newscast Video Data Segment

The newscast information model shown in Fig. 3 depicts the conceptual and structural relationships within newscast video data. For better representation we use object-oriented modeling concepts by treating newscasts as a set of classes.[2] A newscast document class consists of instances of broadcast sessions or a re-composed news document which in turn consists of a number of segmented structural units or news items. The information contained in each news item is stored as object metadata (Section 2.1). Each news item may consist of *1-N* objects (e.g., anchor, Clinton, field footage). An object can be composed of other objects that form a hierarchy of objects or concepts. An object can belong to more than one news item and similarly a news item can belong to more than one document. For example, a train accident can be broadcasted on different sources, or a single instance (one channel) of a news item can belong to different virtual (queried) documents.

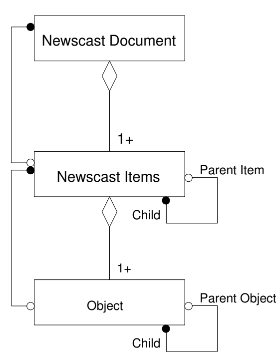

Figure 3: Newscast Video Data Model

The metadata based on the above data model represent objects or concepts in a video

[2]A rectangle in the figure denotes a class, a diamond is a sign of aggregation, a "1+" denotes that there can be one or more objects in an item, an empty circle at the end of a line denotes a single object, and a filled circle denotes multiple objects.

stream (V), an audio stream (A), or combined audio video (AV). In addition to identifying these concepts and the relationships among them, we need to identify the *object ontology* for the concepts represented by the content in the media. The object ontology can be defined based on a user's needs, the application domain, and in our case, on the need for composing and presenting information to a user. In the proposed application these concepts are captured as tokens (text) and are both domain-dependent and domain-independent and stored as metadata (Fig. 4). Based on this concept we define our object ontology.

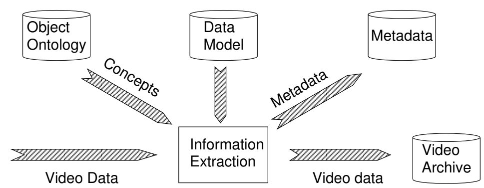

Figure 4: Schematic Representation of the Video Information Extraction Process

2.1 Object Ontology

For automatic customization of news it becomes imperative that we understand how a newscast is composed because the requested information might originate from different sources or time periods. The presentation depends greatly on a user's preference of medium and content. For example, a user can seek a topic comprised of text and images; only associated audio in a specified duration; news about a particular person; or news items from a particular category. The system needs to know what information should be presented, and how. Hence, we create an object ontology to support searching and stream composition. The properties of these objects will aid in information composition. We divide objects into different classes depending on the concepts required in the news composition, as listed in Table 1.

To automate the filtering and the news composition, the information needs to be annotated according to the vocabulary specified. This is ideally achieved by a news provider. The *news segment* is the basic structural unit. The segment is expected to have a coherent linear structure of shots of various lengths and quality. As defined in the ontology, parts of segments can be annotated as relevant introduction, background, current, and enclose, such that when composing at various granularities, we can retain or delete segments (see Section 2.4). The objects discussed as part of ontology are annotated and stored as metadata in a relational database. The database schema aids in fast access to the information, supports information personalization, and information repurposing. A news item can be composed

Table 1: Proposed Object Ontology

Entity	Tangible object part of a video stream.	
Location	Place shown in video.	
Origin	Source where video data are acquired.	
Text	Text can be of the following types:	
	Transcript	Transcript associated with a particular segment of a AV stream.
	Reference	Any additional information (e.g., remarks, critiques, and links).
Graphics	Stills or graphics presented in a newscast.	
Concept	Represent the inferences derived from the presented material. Concepts can be:	
	Entity	Anything that is mentioned in the commentary (e.g., person, thing).
	Location	Associations with certain places and countries that are discussed but not part of the visuals.
	Event & Action	A happenings in a newscast item.
	State	Description of state of an object (e.g., tired, happy, old, and new).
Cinematography	Describe creation-specific information (e.g., video format, title, medium, and playout rate).	
Audio	Audio can be of the following types:	
	Lip Sync	When the audio requires tight synchronization with the video.
	Wild Dialogue	Dialogue that does not sync with a visible speaker [2].
	Voice Over (VO)	When a story uses continuous visuals without showing the speaker.
Footage	Represents the scenes shown. The scenes can belong to one of the following categories:	
	Action footage	Current scenes from the location.
	Archive footage	Material recycled from previous newscasts or any archive material.
	Reenactment	Accurate scenes of situations that are already past or cannot be filmed [2].
Segment	We divide a newscast item into conceptual segments:	
	Headline	Synopsis of the news event.
	Introduction	Anchor introduces the story.
	Background	Gives background information about an event.
	Current	Describes the existing situation.
	Enclose	Contains the current closing lines.
	Complete	A news item which cannot be broken down into previous segments.
	Feature Story	Contains clips that are shot and edited more like magazine pieces.
Category	Classification of news items.	
Reaction	Represents the response of a person or persons to a situation. The response can be acquired by:	
	Interview	One or more people answering formal, structured questions [2].
	Speech	Formal presentation of views without any interaction from a reporter or anchor.
	Comments	Informal interview of people at the scene in the presence of wild sound.

from existing metadata, for presentation and storage, and items can be ordered depending on preferences or priority. In other words, we can form *tours* of the newscast. For example, two users might like to see same information, but in a different order. Then, conceptually, we have two different newscasts or tours even though the content is the same.

2.2 Newscast Database Schema

In Fig. 5 the record type `News Doc` contains general information about a pre-composed newscast provided by a source or composed at run time and stored. The record type `News Item` contains information about each item in the newscast. The record type `Object` contains the metadata about the AV streams. The name of an object, the creation time and date, and the origins make up the composite key for this table. An object can belong to multiple sources (e.g., a clip of Bill Clinton outside the White House). A user might like to retrieve clips of Bill Clinton taken at a certain time or by a certain source. The medium type helps to compose objects from various sources and the popularity field provides information about an object's popularity. We do not store an object stratum (provide access to objects over a temporal span) but the concept of stratification can be easily achieved. The record type `Item Sequencing` defines the tour of the newscast (i.e., the order in which the news items will be presented). The field `Qualifier` is used to represent different tours for the same newscast. Record types `Item Composition` and `Object Composition` define the hierarchy of the news items and the objects. Record types `News-Item Map` and `Item-Object Map` define the news items and objects that are contained in a newscast or in a news item. The record type `Physical Map` represents the metadata of AV and text files. Because we use MPEG 1 compressed video data, to have random seek and playout of video we must use offsets into the video file for starting video playout. In this case we store "Group Start Code" offsets which represent the start of a new "Group-of-Pictures." So we have the field `GSC-File` which represents a file containing offsets for this purpose.

2.3 Proposed Customized News Delivery System

In our existing prototype, we populate the database based on pre-formated newscasts from broadcast television news. A non-overlapping video segment (related juxtaposed shots available from news provider) is the basis of news items composition. When a search is executed, all matching segments are returned without consideration for redundancy of content in multiple items. However, each item is customized based on the available duration and a

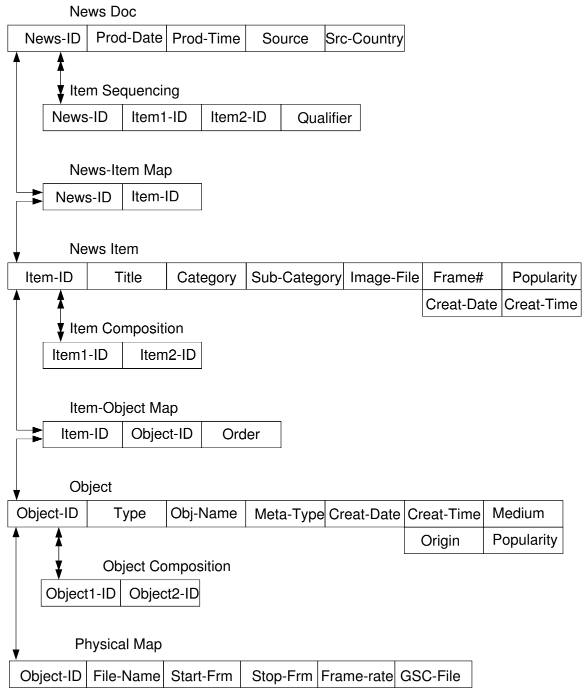

Figure 5: Newscast Application-Specific Network Database Schema

user's preference by adding or deleting the conceptual segments (e.g., introduction, current, background, and enclose) defined in the ontology. We propose to augment the automatic composition of news items by identifying similar video segments by analyzing the similarity among the closed-captioned data. Next we describe the proposed approach for customized news delivery. We assume that we have a user's profile for customization.

1. We assume that video segments received from various sources are already annotated. If not, we are required to annotate these segments and store the information in the record type `Object` in the database. If the metadata received from various sources do not follow the same format then we need to normalize them.

2. We either use the metadata to retrieve matching video segments (location, entity, pre-composed newscast) and compose the clips depending on the associated metadata (intro, current, etc.) or we search the closed-caption data (keyword) to match a query and customize the news item.

3. Based on the ontology we segment the closed-caption data (e.g., intro, current, enclose, feature story, interview, comments). These segments correspond to the segments of the video. We then generate a *segment vector* for each segment. That is, a particular

segment S_i is identified by a collection of terms $Sterm_{ij}$, where $Sterm_{ij}$ represents the weight (importance) of term j assigned to segment i. Based on the measure of similarity between these vectors, we cluster the segments.

4. We create clusters consisting of similar segments. For each cluster we create a *cluster vector* by integrating and normalizing vectors belonging to component segments in each cluster. The cluster vector is used to compare the similarity with the query vector to identify type of information needed.

5. The *user profile*, if any acquired by collecting explicit user preferences and the *query* are converted into vectors similar to segment vectors. The keywords within these vectors can have different weights (intra-weights) between 0 and 1 (e.g., if a location concept vector contains keywords like India, Asia, and USA, and India might have the highest weight and USA the lowest weight). A vector can also carry a weight (inter-weights) (e.g., a location concept vector will carry more weight than the event concept vector).

6. The weights in the *query vector* are normalized by the *user profile vector*. Such that the terms in query vector are weighed appropriately according to a user's preferences.

7. The cosine measure is used to measure the similarity or the cosine of the angle between a query vector and a cluster vector as follows:

   $Cosine(Cluster_i, Query_j) =$

   $$\frac{\sum_{k=1}^{t}(Cterm_{ik} * Qterm_{jk})}{\sqrt{\sum_{k=1}^{t}(Cterm_{ik})^2 * (Qterm_{jk})^2}}$$

8. Next we compose video clips associated with retrieved segments depending on a user's preference and the duration of presentation.

The following are the rules which are followed during information composition and filtering of chosen segments from each cluster:

- The age or the creation time of the segment plays an important part in the selection. A recent segment should have a higher probability of being included in the composition of news item. However, aged news or *archived news* can be used to provide background information as filler.

- The degree of similarity between various segments in a cluster is proposed as another means of selecting a segment and maintaining continuity. We create *sub-clusters* (e.g.,

Table 2: Example Metadata

Obj-ID	Type	Name	Metatype	Time	Date	Medium	Origin	Popularity	Start Frame	Stop Frame	Item -No.
O01	Person	Lyle	Anchor	07:30:00	06-27-96	AV	CNN	1	0	165	
O02	Location	Studio	Actual	07:30:00	06-27-96	V	CNN	1	0	165	
O03	Audio	Lip Sync	Anchor	07:30:00	06-27-96	A	CNN	1	0	165	
O04	Graphics	Map	Durham	Null	Null	V	Library	1	166	837	
O05	Audio	VO	Anchor	07:32:00	06-27-96	A	CNN	1	166	837	
O06	Person	Clinton	President	19:00:00	06-26-96	AV	Reporter1	0.7	838	2314	
O07	Reaction	Speech	Concept	19:00:00	06-26-96	A	Reporter1	0.7	838	2314	
O08	Location	White House	Actual	19:00:00	06-26-96	V	Reporter1	1	838	2314	
O09	Audio	Lip Sync	Clinton	19:00:00	06-27-96	A	Reporter1	1	838	2314	Ist
O10	Person	Lyle	Anchor	07:35:00	06-27-96	A	CNN	1	2315	2747	
O11	Audio	Lip Sync	Anchor	07:35:00	06-27-96	A	CNN	1	2315	2747	
O12	Location	Studio	Actual	07:35:00	06-27-96	V	CNN	1	2315	2747	
O13	Person	Reporter2	Reporter2	12:15:00	06-27-96	AV	Reporter2	0.4	2748	2928	
O14	Bob	Person	Senator	12:15:00	06-27-96	A	Reporter2	0.4	2748	3267	
O15	Reaction	Comment	Concept	12:15:00	06-27-96	A	Reporter2	0.4	2748	3267	
O16	Segment	Bombing	Intro	07:40:00	06-27-96	A	CNN	1	0	837	
O17	Location	Dhahran	Concept	07:41:00	06-27-96	AV	CNN	1	0	3267	
O18	Location	Saudi Arabia	Concept	07:41:00	06-27-96	AV	CNN	1	0	3267	
O09	Text	Transcript	Null	07:30:00	06-27-96	T	CNN	1	0	3267	
O20	Event	Bombing	Concept	07:41:00	06-27-96	AV	CNN	1	0	3267	

intro and current) within each cluster and choose only one segment from each cluster. We use a random element in our selection from each sub-cluster to avoid the possibility of the same segment being selected each time.

- The duration of the playout also plays an important factor in news item composition. For example, if the playout duration is shorter as compared to the data available then "current" sections of the segments will take priority over others.
- The composition of the news stream should yield ordering of items in the delivery as a function of the user's interest. Items of higher interest should be played out first.
- The detail contained in news items should also be a function of a user's interest. If playout is time-constrained then the higher interest news items should get more time than others.

2.4 Stream Composition Example

To illustrate our metadata organization, consider the following example newscast by CNN. This newscast is annotated in Table 2.

Here object "O_{08}," represents a place "White House" that is shown in the video stream. Object "O_{15}" represents a conceptual location. "Dhahran" is mentioned by the anchor person but is not actually shown in the video stream. An object is mapped to physical streams by the `Object-ID`. The `Physical Map` contains the segment's start and stop frame with which an object is associated. Objects O_{01} to O_{20} belong to the news item "Bombing in Dhahran." The last three columns in the table are not part of `Object` metadata; we have inserted them for better visualization of the news items.

The first few sections of the news item consist of an anchor person relating the event of bombing and the rest of the segments consist of the reactions of Clinton and Dole. If the shots from the field are available, the news item can be made much more interesting by inserting these shots. One approach is to compose the anchor person's audio containing the introduction with these shots. Therefore instead of seeing a map of Dhahran we would see the actual event. If duration of presentation does not allow complete playout of an item, we can play the scenes of the event with voice-over of anchor person introducing the event i.e., replacing or adding voice to the existing recording [11]. This process of composition is potentially more informative and concise.

In previous sections we proposed the model for delivering customized news. The following are the factors used to judge the quality of a news video stream. Of course, in our scheme, the quality of the results will depend on the quality and application of the collected metadata.

- **Segment non-repetition** As we combine video data received from various sources, the news items should not contain redundant information. Let ϑ be a metric of similarity of news segment vectors. To evaluate a news item for redundancy of information, we compare each news segment (S_i, S_j) with the others. If the *similarity threshold* is below the ϑ then the segments belong to different clusters (C_k and C_{k+1}). The similarity threshold is further increased and the segments within each cluster and further grouped into sub-clusters. This can be used to measure success in composition.

$$\begin{aligned} Cosine(S_i, S_j) &\geq \vartheta, \text{ then } (S_i, S_j) \in C_k \\ &< \vartheta, \text{ then } S_i \in C_k, \;\; S_j \in C_k + 1 \end{aligned}$$

- **Thematic flow** The continuity in information flow should be maintained. Let δ be a metric for continuity of flow of information in a news item, or *thematic threshold*. Every consecutive pair of sub-cluster vectors (SC_{ij}, SC_{ik}) are compared for similarity, and if the thematic threshold is greater than δ then there is a deviation in the flow of information and segments belonging to these sub-clusters are not placed to each other. This can be used to determine success in achieving a gradual change in the theme of the news item.

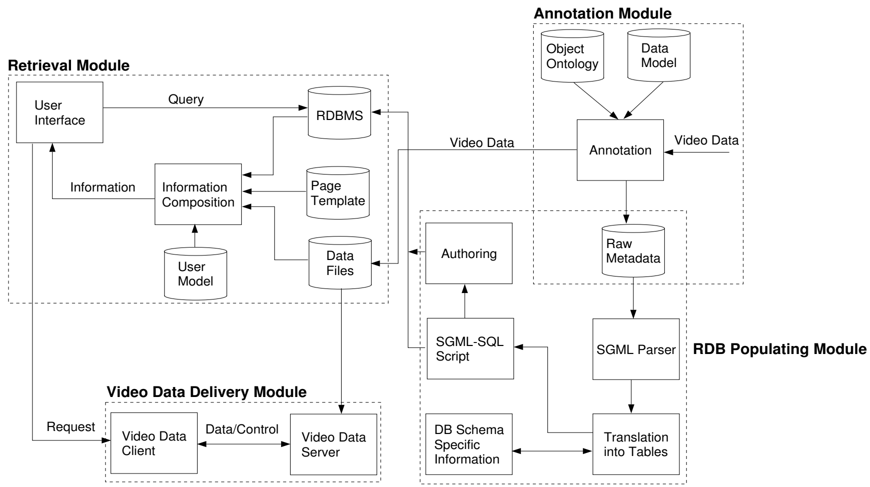

Figure 6: Data Flow Diagram for the Prototype

- **Segment matching** The news delivery should contain news items which match the user need. Let Δ be a metric for relative measure of similarity (*cut-off threshold*) between the first segment of a news item and the last segment of the news item. We use this to determine success in limiting the scope and duration of a news item.

We describe the system design next.

3 A News Video Composition and Delivery System

The overall functional modules and data flow in the system are shown in Fig. 6.

The annotation module consists of processes for transforming information contained in video data into metadata. We store the information model as a document type definition (DTD). The process of annotation involves loading a DTD into Vane, an annotation tool for video data [5]. This tool uses structural information to annotate the data by breaking up video data into shots, scenes, and sequences. Data forms associated with each structural unit are used to annotate the content. Captured closed-caption sections are indexed using SMART [4]. These indices are searched when a keyword query is applied to the system. The information about the type of segment (e.g., introduction, question and answer) that the closed-caption section belongs to is stored as metadata. The metadata is stored as SGML-compliant markup/metadata.

After validation of SGML markup we use the `SGML2SQL` converter to map the information in the raw metadata to the database schema discussed in Section 2.2. This process populates the appropriate tables and fields. For example, a scene is mapped to a news item and all associated shots and content information are mapped to objects. The start and stop frames associated with these objects are mapped to the table representing the physical data schema.

A World Wide Web interface is used for making queries to the news information database while HTTP provides communication between the interface and the data server. The system composes information according to the templates provided. Finally, because HTTPD is not suitable for streaming data, a separate communication channel is established between the media client and the media server to deliver video.

All of the above subsystems interact to provide a user with desired information from the video database. Whenever a selection is made, the video segments are composed according to the order specified in the `Item-Object Map`. Based on a users request, a range of information from a summary to a complete item can be provided.

4 Summary

In this paper we have defined concepts and techniques which help in information filtering, retrieval, and automatic composition of a news video. We have proposed a complete system for news information retrieval. This involves formulating the concepts on which information retrieval and composition is based. We have presented a data model defining the relationships among the different concepts and a database schema appropriate for managing metadata. We have proposed concepts and their attributes to assist in automatic customization of a newscast.

Our current system does not identify segments automatically. However, we are actively pursuing this goal. To achieve this, we require annotation of collected/digitized news video. For this purpose we have a facility to digitize analog video data as MPEG in real-time. These data are stored on a server capable of holding in excess of 200 hours of full-frame rate video. Based on the described infrastructure, we intend to evaluate the proposed automatic news composition system.

5 Acknowledgement

We extend thanks to William Klippgen for his investigation and insight into the problem of metadata collection from news video data.

References

[1] N.J. Belkin and W.B. Bruce, "Information Filtering and Information Retrieval: Two Sides of the Same Coin?" *Communications of the ACM*, Vol. 35, No. 12, pp. 29-38, 1995.

[2] M. Brabiger, *Directing the Documentary*, Focal Press, Boston, 1992.

[3] M.G. Brown, J.T. Foote, G.J.F. Jones, K.S. Jones, and S.J. Young, "Automatic Content-Based Retrieval of Broadcast News," *Proc. 3rd ACM Multimedia '95*, pp. 35-43, 1995.

[4] C. Buckley, "Implementation of the SMART Information Retrieval System," Computer Science Department, Cornell University, No. TR85-686, 1985.

[5] M. Carrer, L. Ligresti, G. Ahanger, and T.D.C. Little, "An Annotation Engine for Supporting Video Database," to appear in *Multimedia Tools and Applications*.

[6] P.R. Chesnais, M.J. Mucklo, and J.A. Sheena, "The Fishwrap Personalized News System," *Proc. IEEE 2nd Intl. Workshop on Community Networking Integrating Multimedia Services to the Home*, 1995.

[7] Excite Inc., "NewsTracker," *http://nt.excite.com*.

[8] P.W. Foltz and S.T. Dumais, "Personalized Information Delivery: An Analysis of Information Filtering Methods," *Communications of the ACM*, Vol. 35, No. 12, pp. 51-60, 1992.

[9] T. Kamba, K. Bharat and M.C. Albers, "An Interactive, Personalized, Newspaper on the WWW," *Proc. 4th Intl. World Wide Web Conference*, 1995.

[10] G. Miller, G. Baber, and M. Gilliland, "News On-Demand for Multimedia Networks," *Proc. 1st ACM Multimedia '93*, pp. 383-392, 1993.

[11] R.B. Musburger, *Electronic News Gathering*, Focal Press, Boston, 1991.

[12] H.J. Zhang and S.W. Smoliar, "Developing Power Tools for Video Indexing and Retrieval," *Proc. IS&T/SPIE, Storage and Retrieval for Image and Video Databases II*, pp. 140-149, 1994.