BOSTON UNIVERSITY

COLLEGE OF ENGINEERING

Dissertation

# TECHNIQUES FOR AUTOMATIC DIGITAL VIDEO COMPOSITION

by

GULRUKH AHANGER

B.E., Kashmir University, 1987

M.S., Boston University, 1993

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
1999

Approved by

First Reader: ______________________________

Thomas D.C. Little, Ph.D.
Associate Professor of Electrical and Computer
Engineering
Boston University

Second Reader: ______________________________

Stanley Sclaroff, Ph.D.
Assistant Professor of Computer Science
Boston University

Third Reader: ______________________________

Richard F. Vidale, Ph.D.
Professor of Electrical and Computer
Engineering
Boston University

Fourth Reader: ______________________________

John F. Buford, Ph.D.
Associate Professor of Computer Science
University of Massachusetts Lowell

## Dedication

*To My Mother*

Whose memories shall always be with me

and

*To My Father*

For his love, sacrifice, encouragement, and patience

and

*To Kashmir*

Where I was born and the thoughts of which sustain me

# Acknowledgments

I owe many debts of scholarship and support to many people.

First and foremost, I would like to thank my advisor, Prof. Thomas D.C. Little, without whose inspiration, confidence and support, this dissertation would have remained merely a dream. Ever since I joined the Multimedia Communications Laboratory, Prof. Little has always been there to provide guidance, friendship, and encouragement, and I just cannot thank him enough.

I benefited from the advice and guidance of many other faculty members. In particular, Professors Stanley Sclaroff, Richard Vidale and John Buford were especially generous with their time, support and encouragement, of which this dissertation has been the primary beneficiary. It would be remiss of me not to express my gratitude for Prof. Brackett and Prof. Ruane, both of whom supported me through my initial years at the Boston University. In general, the entire faculty and staff of the Electrical and Computer Engineering Department has been very supportive, giving, and helpful, and I remain grateful to all of them for making the process of my education so enjoyable.

I owe a lot to my colleagues at MCL, particularly Anand Krishnamurthy, Dinesha Venkatesh, Rajesh Krishnan, William Klippgen, Prithwish Basu, Ashok Narayanan, Karthikeyan Srinivasan, Leonardo Ligresti, Leslie Kuczynski, Marco Carrer, Jacob Begin, James Martin, and John Casebolt. Besides specific advice and support, my colleagues helped me maintain a keen sense of humor, balance, and equilibrium through all those never-ending days, many of them punctuated by frustration and pain!

The Palayoors and the Chawlas were family to me, and I am grateful to them for allowing me to so freely encroach on their time and affection. I am grateful to Lexa Iantuono for always being there whenever I needed help, laughter, and cheering up. Many other friends go unmentioned here, but just as they have indulged me in the past, so shall they forgive me for maintaining their anonymity.

Last, I thank my family for all that they have given me. This dissertation is a tribute to their love, patience, and understanding.

TECHNIQUES FOR AUTOMATIC DIGITAL VIDEO COMPOSITION

(Order No. )

GULRUKH AHANGER

Boston University, College of Engineering, 1999

Major Professor: Thomas D.C. Little,
Associate Professor of Electrical and Computer
Engineering

## Abstract

Recent developments in digital technology have enabled a class of video-based applications that were not previously viable. However, digital video production systems face the challenge of accessing the inherently linear and time-dependent media of audio and video, and providing effective means of composing them into a cohesive piece for presentation. Moreover, there are no appropriate metrics that allow for assessment of the quality of an automatically-composed video piece. Techniques presently available are limited in scope, and do not account for all the features of a composition. This dissertation presents metrics that evaluate the quality of a video composition. In addition, it proposes techniques for automatic composition of video presentations as well as improvements in access to digital video data.

Yet another challenge faced by video production systems is the customization of the presentation to suit user profiles. For instance, certain elements of video compositions, such as violence and indecent exposure, are undesirable for some audiences. Also, playout time of a composition can be longer than specified by the user. In such cases, not only would some of the data need to be dropped, the integrity and cohesiveness of the composition must also be maintained. This dissertation presents techniques for maintaining cohesiveness of a composition under playout time constraints.

Using automatic composition techniques proposed in the dissertation, a video piece is produced. The quality of a manually-produced broadcast news video composition is evaluated using the metrics, yielding reference values of video composition quality. The same metrics are used to evaluate the quality of the video piece produced using automatic composition techniques. A comparison of the two indicates that the quality of the automatic composition is very similar to that of the manually produced video composition; in some cases it is superior. These results also verify the assumptions on which the automatic composition techniques are based.

In addition to the metrics and the video composition techniques, a methodology to improve the recall and retrieval of a digital video production system is proposed. Two search techniques are used: transitive search and union-based search. The proposed methodology is implemented as part of a digital news video production system. An analysis of the performance of the methodology shows an increase in recall by 23% when the transitive search technique is used, and an increase of 48% when the union-based search technique is used, as compared to a keyword-based search technique.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Several factors influence recent advances in digital video-based communication systems. These include developments in digital technology, federal regulations, and industry-based efforts. Developments in digital technology include higher network bandwidth, streaming-enabled data transfer protocols, large-scale storage servers, digital video capturing equipment, video compression techniques, and high-end multimedia-enabled workstations, all of which are contributing significantly to the development of digital video-based communication systems. In a bid to expedite the move from analog to digital technology, the federal government, through the FCC, is also supporting regulation to convert television stations to digital broadcast [71]. Finally, recent industry-based developments, such as the integration of Web-based technology with television [26, 71], and Microsoft's Broadcast Architecture for Windows that claims to allow user choice of content from varied sources, are also supporting and promoting the development of digital based communication systems. Indeed, the combined effect of all these diverse efforts of technology, state, and industry portends well for the imminent development of digital video-based communication systems.

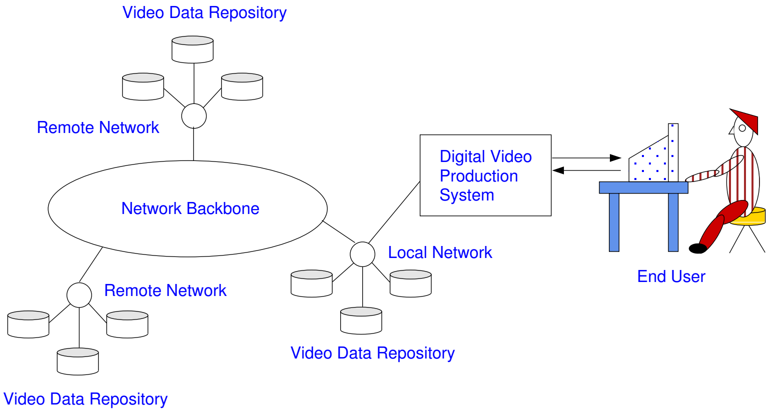


Figure 1.1: A Video-Based Communication System

Digital video-based communication systems (Fig. 1.1) are expected to support such video-based applications as newscasting, sportscasting, and distance learning. Since such applications provide automatic access to linear and time-dependent video and audio media, use of digital video-based technology can potentially open up interesting editorial opportunities both within (e.g., a scene of a movie) and across multiple instances of the medium (many movies). Thus, a video that has been used to create a movie or a news story need not be confined to a single rendering; it can also be used in multiple contexts without involving an extensive re-production process. In other words, once an access is achieved, additional manipulation of the video media is possible to produce a narrative, or a series of episodes collected as a chain in a storyline [19].

In conventional production systems, a human decides which video segments should be used and how a narrative should be assembled. The production of a complete video piece involves three phases: pre-production, production, and post-production. In the pre-production phase, before creating a video, an underlying concept or storyline is developed that serves as a guide for production efforts. For example, in electronic news gathering (radio or TV broadcast), a storyline is developed based on a current event or other cultural, social, political, or experimental curiosity [63]. Shots that create a beginning, middle, and end of a story are formalized conceptually. Once these items are determined, they are scripted. Thus, a script contains detailed instructions of how and what is to be shot and serves to minimize effort in the shooting process. In electronic news gathering (ENG), a script can span many news items and can consist of many pages of text.

The next phase in conventional production systems is the production phase, which involves the shooting of raw video footage. The location is prepared, equipment is set up, and lights are arranged. A shot is composed while taking care of balance and symmetry. Then the actual film shooting occurs and information about the shot is logged. The process is repeated until all desired footage is complete, including shots recorded to provide continuity between core pieces.

The final phase is the post-production phase, in which the raw video is manipulated and prepared for distribution. In ENG, or documentary-making, a post-production script is prepared that describes all the relevant information. The script can be written before or during editing. This script is read (e.g., by an anchor person) in conjunction with the edited video. Usually the pre-recorded video shots are delivered to an editing point where shots or frames are cut and composed for the final piece.

Of the three stages involved in conventional video production, digital-based video communication systems can fully automate only the post-production stage. Presently, some degree of automation in the post-production phase is provided by a variety of personal-computer-based solutions that aid the human operator. Segments can be easily recorded and manipulated with special effects such as wipes, dissolves, fading in/out, distorting, and embossing. Digital video editing packages such as Adobe Premier, Kohesion, and MediaStudio provide robust tools for commercial and professional use [86]. In addition to functions for selection, transitions, and trimming, operations including ripple and rolling edits, multiple-track selection, jog, shuttle, and play enable large amounts of footage to be quickly edited.

However, in order to support dynamic video composition and delivery, we require automatic selection and manipulation of video segments from an archive, which the digital-based video technology can provide. Two types of manipulation of video media are possible: first, we can shuffle segments in a composition based on some user-defined criteria; and second, in addition to shuffling segments, we can dynamically compose a segment (i.e., compose various content objects to create a segment) [59]. In our work we only consider segment manipulation to compose video and not the creation of a segment itself.

Before we can realize our vision of dynamic and automatic composition of video, at least three issues need to be resolved: information requirements, information extraction methodology, and video composition techniques. A semi-automatic system requires information about content for editing and composing a video piece. The system also requires techniques for composition. Therefore, identifying the information sufficient for editing and composition, determining how the information should be extracted, and creating techniques for cohesive video composition are issues that need to be addressed in order to develop a digital video production system.

A typical digital video production system (DVPS) requires a video data model and ontology, a mechanism for information extraction, a mechanism for interactive query, a user model, and a mechanism to compose and customize data. (Fig. 1.2 illustrates a functional view of a DVPS). We summarize these components as follows:

**Data Model & Ontology:** An ontology establishes the domain-specific concepts required and the relationships among the concepts [53]. The concepts represent both content as well as structural information [22]. A data model represents the concepts/information and the relationships [30, 88].

**Information Extraction:** Based on the data model/ontology, concepts/information are extracted and stored as metadata. Information can be extracted manually, automatically, or by combination of the two [7]. Automatic extraction depends heavily on image processing tools [10, 13, 16, 37, 38,

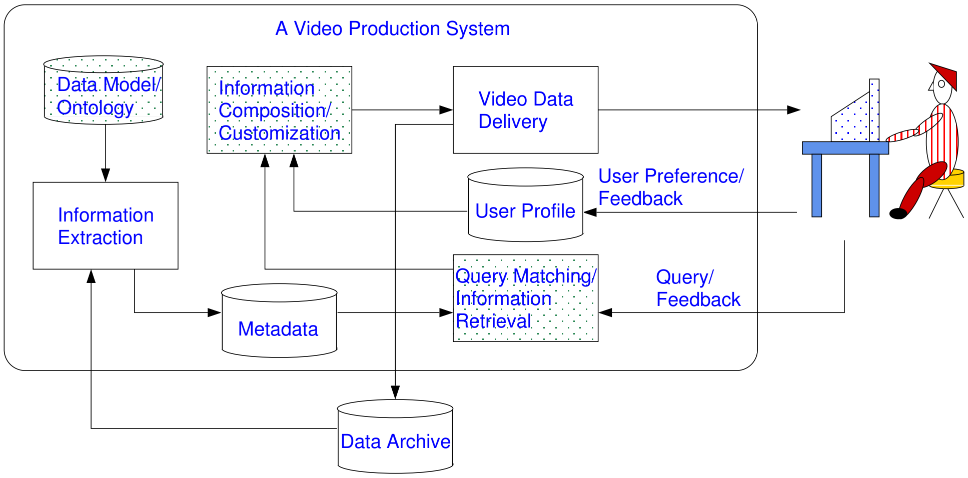


Figure 1.2: Functional View of a Digital Video Production System

39, 43, 49, 65, 67, 73, 94]. Information within unstructured data such as video is easily identified by human observation; however, few attributes can be identified by a machine. Therefore, we are more dependent on hybrid extraction techniques.

**User profile:** A user profile represents information about user behavior and preferences [47]. *Canonical* and *descriptive* are the two main classes of user models. The canonical model requires a formal encoding of a cognitive (semantic) user model [48, 57]. These models are hard to acquire and their complexity hides the represented semantics from the user. Descriptive user models can be automatically created by observing user behavior [70]. Their content is a mapping from previous document accesses and does not require any semantic processing. However, a large number of observations is needed to be able to draw high quality conclusions.

**Interactive Query:** Interactive query is a process of formulating a query and matching the query with the available metadata [8]. Several techniques have been proposed for retrieval of video data using visual methods, most of which fall within the three categories of query by example (QBE), iconic query (IQ), and keyword-based query [15, 18, 23, 33, 35, 38, 41, 44, 69, 89, 90, 91]. QBE queries are formulated using sample images, rough sketches, or component feature of an image (outline of objects, color, texture, shape, layout). These queries make extensive use of image processing and pattern recognition techniques. In keyword-based and iconic query extracted concepts are used for matching the query.

**Composition and Customization of Video Data:** We define video composition as a process of assembling video segments into a logical and thematically correct depiction of a storyline. Video can be composed using visual ranked-based, text rank-based, temporal-based, and rules-based techniques [3, 31, 64, 72, 78, 92]. Visual and text ranked-based composition techniques utilize weights assigned to the concepts within visuals and audio data for assembly. In a temporal composition data are assembled based on temporal relations of concepts within their contents (e.g., "retrieve a video piece in which Blair is waving before he presents a speech"). In rules-based composition additional content-based and structure-based constraints are imposed on a composition (e.g., topics should be introduced before discussing them in detail).

Based on a user profile or system requirements a composition can be tailored or customized. We divide customization into three categories: content-based, structure-based, and time-based. In content-based customization only required information is provided and the rest is filtered [64, 72]. In structure-based customization only requested parts of a structure are composed (e.g., headlines of the latest news). Time-based customization deals either with the relative position of segments on a timeline (e.g., "retrieve news sports, and then stocks") or the playout duration (e.g., "recap news for two minutes") [3, 72].

In this dissertation, we address the issues related to composition and customization techniques. Existing composition techniques are not adequate for producing a narrative. These techniques are ranked-based and content-rule-based; the temporal dependency of video data and their domain-specific structure are not considered, hence, utility of these techniques is limited and does not always result in a correct narrative (e.g., retrieve all information about the gondola accident in Italy from start to end). Furthermore, a video presentation can be comprised of single composition, or, as seen in Fig. 1.3, a presentation can consist of multiple pieces of composed video, (e.g., a newscast presentation is comprised of individual news items). Therefore, we require techniques that help create a presentation with multiple compositions. As part of this work, we present a query and selection technique that retrieves data from a corpus (universal set) and forms an individual candidate set for composition.

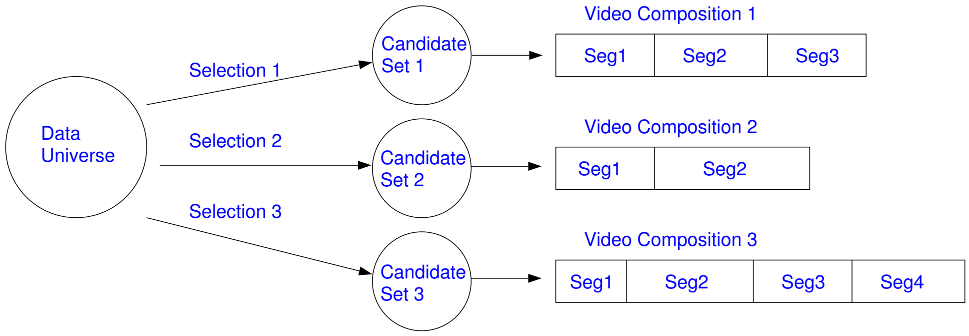


Figure 1.3: Process for Composition of a Video Item

Besides accomplishing automatic composition, the quality of these compositions must be compared with man-made compositions. However, we are not currently aware of any metrics that can evaluate a composition.

This dissertation attempts to fill this gap. In particular, we we focus on *metrics* for evaluation of a composition and *techniques* for composition and customization of digital video. To this end, we directed an investigation into the features of manually composed video pieces. To support the composition techniques, we implemented a prototype DVPS for newscasts called Canvass (Customized Access to News Video Archive Storage System). During implementation, we also observed the information semantics within related video data. Based on these observations we propose a hybrid retrieval technique presented as part of a system implementation.

## 1.1 Contributions

The following are the main contributions of this work:

- We propose a methodology for composition of video pieces. The methodology includes *instance-based* and *period-based*. “Retrieve the latest news” is an example of an instance-based composition and “retrieve all the news about the Pope’s visit to Cuba” is an example of a period-based composition. The instance-based narrative is composed along a storyline while maintaining the structure of the domain. The period-based narrative is composed along a storyline, the facts are presented in the order they developed, and the structure of the domain is maintained.

- We propose a set of metrics that best reflect characteristics of a composition. The characteristics include amount of information presented to a user in a composition, thematic flow in a composition, temporal flow in a composition, content progression in a composition, period span coverage by a composition, and domain-specific structure of a composition.

- We propose a novel four-step hybrid approach for retrieval and composition of video that improves the recall of related data. The information tends to vary among related segments. For example, it is common in broadcast news items that once an event is introduced, in subsequent scenes the critical keywords are alluded to but not specifically mentioned. Related segments will not have common critical keywords, but, scenes may share other keywords. Not all directly related segments are necessarily retrieved if a search is made on a person's name. Similarly, related video segments can have different visuals. The proposed approach overcomes these limitation of video data semantics.

In brief, we propose metrics that comprehensively represent a composition and quantify the quality of a video composition, composition and customization techniques for video data that are based on content, time, and structure. In addition, we propose a hybrid technique for retrieval of related video segments. The proposed techniques are demonstrated and evaluated using newscast video data.

## 1.2 Organization of the Dissertation

The remainder of this dissertation is organized as follows: In Chapter 2 we discuss existing techniques for evaluation and composition of digital video composition and their limitations. In Chapter 3 we propose a set of metrics that are used to evaluate an automatic digital video composition and evaluate the quality of broadcast news. In Chapter 4 we propose techniques for composition and customization of a digital video. In Chapter 5 we evaluate a composition achieved by the proposed techniques and compare the quality with that of broadcast news. In Chapter 6 we discuss the concepts behind the implementation of a news DVPS. In Chapter 7 we discuss the system architecture and implementation of a news DVPS. The news DVPS is implemented to support the composition and customization techniques. In Chapter 8 conclusions and directions for future work are presented.

# Chapter 2

# Background and Related Work

## Synopsis

Several different metrics and techniques are used for evaluation, information retrieval, and presentation. In this chapter, we discuss these metrics and techniques, and highlight their limitations in video composition. In particular, we describe the existing metrics that are used for evaluation of both text and discrete multimedia data (e.g., images and video segments) retrieval systems. We also describe the existing segment-based, pre-assembled, and dynamically-assembled video presentation techniques as well as the features (e.g., theme, time, and structure) that should be considered during dynamic assembly of video data. Finally, we describe the existing customization techniques including content-based, time-based, structure-based, and cost-based video. In each case, we highlight the limitations of the metrics and techniques in video composition, and identify specific areas of deficiency in evaluation and composition of a video piece.

## 2.1 Metrics for Performance Evaluation of Information Retrieval Systems

Information retrieval (IR) systems support access to large data corpora including, text, images, graphics, audio, and video data. Information retrieval systems mainly use metrics proposed by Salton [79] to evaluate data retrieval performance. The metrics measure *recall* (R) and *precision* (P) of an information retrieval system. These metrics remain valid for IR; however, these metrics are oriented towards Boolean evaluation (i.e., a retrieved object either matches a query or it does not) and do not consider the degree of similarity between the user criteria and retrieved results. Recall measures the ability of the system to retrieve all relevant data. Precision measures the ability of the system to present relevant data.

$$R = \frac{number\ of\ items\ retrieved\ and\ relevant}{total\ items\ relevant\ in\ collection}$$
$$P = \frac{number\ of\ items\ retrieved\ and\ relevant}{total\ retrieved}$$

In addition to the above metrics, ranked evaluation metrics are also used to measure retrieval performance. In this case, a retrieved object does not exactly match the query but has a degree of similarity. Narasimhalu et al. [66] have proposed metrics for retrieval of multimedia objects. Their metrics measure the *rank, order, spread,* and *displacement* of retrieved objects. These metrics are summarized below.

*Order:* Order quantifies the ability to sequence data items in the retrieved set. In the example below the system retrieves data in an incorrect order.

Example 1.

Correct response: $o_1, o_2, o_3, o_4, ....$

Actual response: $o_2, o_4, o_3, o_1, ...$

*Rank:* Rank measures the degree of relevancy of the retrieved set to the query. In the example below the rank of individual objects in the retrieved set is less than the actual rank.

Example 2.

Correct response: $o_1, o_2, o_3, o_4, ....$

Actual response: $o_7, o_2, o_4, o_3, o_1, ...$

*Spread:* Spread measures the shift in the position of a data object in the retrieved set as compared to the correct position. This is illustrated in Example 3.

Example 3.

Correct response: $o_1, o_6, o_2, o_3, o_4, ....$

Actual response: $o_1, o_2, o_8, o_9, o_3, o_4, ...$

*Displacement:* Displacement measures the position of a data object in the retrieved set as compared to its correct position. This is illustrated in Example 4.

Example 4.

Correct response: $o_1, o_2, o_3, o_4, ....$

Actual response: $o_1, o_2, o_4, o_3, ...$

After the degree of similarity has been established in retrieval, the ordering becomes trivial as segments are re-ordered to create a narrative. Spread and displacement metrics are another means of specifying the performance of recall and ranking of the system, and provide little added information about the performance of the system.

The ranked-based/approximate/fuzzy retrieval systems use *similarity* assessment techniques to match data with a query. Most of these techniques are based on distance measure in some perceptual space. The most commonly used measurement is the Euclidean metric [14].

$$d(o_1, o_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

where $o_1$ and $o_2$ are objects to be measured and $x$ and $y$ represent their feature space.

Salton proposed a *cosine* similarity metric for measuring the similarity between two document vectors in the multidimesion feature space $t$;

$$\text{cosine}(\vec{doc_i}, \vec{doc_j}) = \frac{\sum_{k=1}^{t}(Term_{ik} \times Term_{jk})}{\sqrt{\sum_{k=1}^{t}(Term_{ik})^2 \times \sum_{k=1}^{t}(Term_{jk})^2}} \tag{2.1}$$

This metric measures the cosine of the angle between the two documents. The numerator of the cosine metric gives the sum of matching terms (or term weights) between the two documents. The denominator is a product of the lengths of the two documents and acts as a normalizing factor. If the evaluation is not binary, each term in a vector is represented by a weight evaluated by various weighting schemes. Most commonly used schemes are *term frequency* and *inverse document frequency.* Term frequency is based on the notion that the terms that occur more frequently have some relation to the content of the texts. Term frequency makes no distinction between the terms that occur in every document of a corpus and terms that occur in only a few. Inverse document frequency calculates the weight of a term based on the concept that the importance of a term increases with its frequency in a document but decreases with the number of documents ($DocFreq$) for which its is assigned. The weight of term $k$ in document $i$ is:

$$weight_{ik} = \frac{Freq_{ik}}{DocFreq_k}. \tag{2.2}$$

Many variations of the inverse document frequency weighting schemes are used [85] to calculate term weight. For example,

$$weight_{ik} = Freq_{ik} \times (log_2 \frac{n}{DocFreq_k} + 1)$$

Transcripts associated with video data can be indexed (weights assigned) and represent a keyword vector with which a query is matched. A similarity value can be assigned between a transcript vector and a query. Brown et al. [20] use transcript data to deliver news data. In the Informedia project [91] transcripts are used to extract video segments for browsing. Wachman [90] correlates transcripts with the scripts of situation comedies. The script discloses who says what and the transcript specifies the precise position in the video data where it was said; hence, video can be automatically indexed with characters, shots, and scenes.

In addition, concepts/information contained within visuals can be used for retrieval. Based on the concept vectors, cosine metrics can be used to evaluate similarity among video segments.

Though many of the previously mentioned metrics can be used to evaluate the performance of a retrieval system, they are not useful in evaluation of a video composition. However, a DVPS is not only required to retrieve data but also to achieve a composition. A storyline, or a theme, must be maintained in a composed piece. Therefore, a new set of metrics is required to quantify a DVPS's composition performance.

Next, we discuss segment-based, pre-assembled, and dynamically assembled presentation techniques for video data. Since our focus is on dynamically assembled video data, we review the existing work in this domain and discuss its limitations.

## 2.2 Video Data Composition Techniques

There are three types of techniques used to present video data. First, video data can be presented as discrete segments with no established relationship among the segments. Second, video data can be pre-assembled, (e.g., video segments assembled for a particular movie delivery). Lastly, enough information can be made available to the system to assemble data on-the-fly for delivery. These presentation techniques are summarized below.

**Segment-based presentation:** This is a trivial presentation of information, in which information is presented in the form of discrete video segments (Fig. 2.1). Discrete segments are retrieved from a heap (e.g., "retrieve all the clips that have Clinton playing Saxophone"), and presented to the user. The relationships between various segments (clips) are not evaluated.

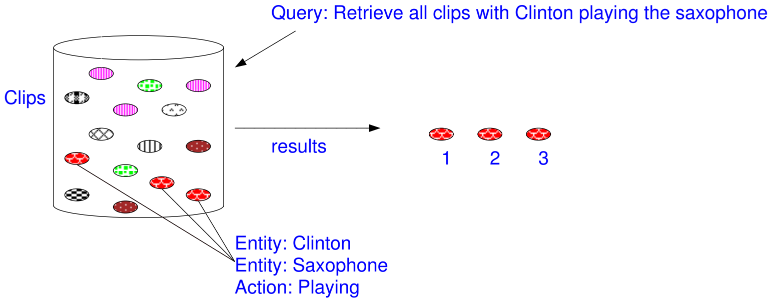


Figure 2.1: An Example of a Segment-Based Retrieval

**Pre-assembled presentation:** In this presentation technique, the content is pre-orchestrated [5]. In other words, the information about segments composing a presentation and their order are stored as metadata. Fig. 2.2 shows that the information about topics and the order they need to be presented are stored as metadata. Depending on a query, respective paths are traversed to retrieve information. For pre-assembled video data there is little freedom of customization or reorganization.

**Dynamically-assembled presentation:** The sequencing of video segments in a presentation or a narrative is achieved on-the-fly (Fig. 2.4). In

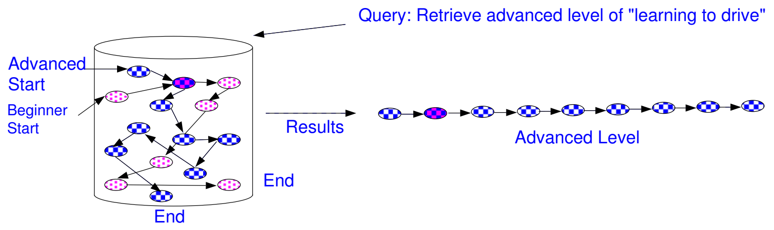


Figure 2.2: An Example of a Pre-Assembled Retrieval for a Lesson Plan

a dynamic-assembly, instead of having information about a complete narrative, information about the content in a narrative is stored. The information within individual video clips is used to compose and customize a narrative.

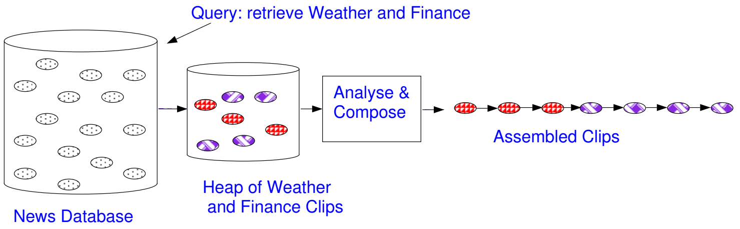


Figure 2.3: An Example of a Dynamically-Assembled Retrieval for a Newscast

Once content is selected for assembly and a chain is formed, the content is mapped to the timeline called *spatio-temporal* mapping. The spatio-temporal mapping of the structures can belong to one of the three scenarios described below.

1. Structures in one creation time reference are mapped to a playout timeline as shown in Fig. 2.4. Video clips containing desired concepts or

information are excerpted from a recorded storage medium (e.g., tape, digital file) and are ordered on a timeline for playout. The clips are arranged in the order they are created.

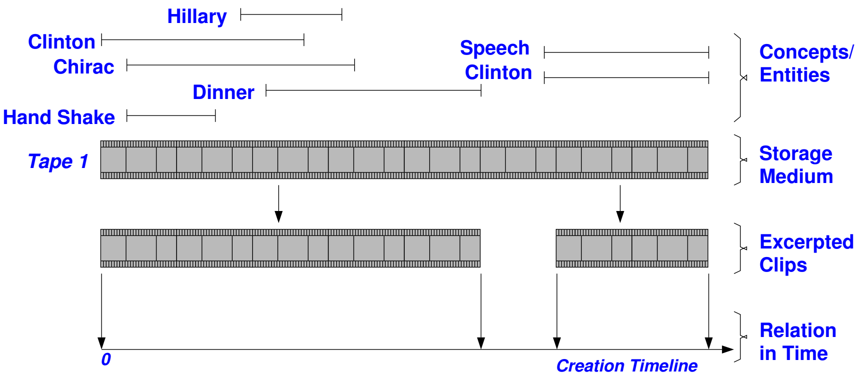


Figure 2.4: Spatio-Temporal Mapping in One Time Reference

2. Structures across multiple references (tapes) of the creation timeline are mapped to a playout timeline. Multiple references can overlap in time, that is, more than one reference can have information from the same period on the creation timeline. In Fig. 2.5 we show the references in two media overlapping.

3. Structures in creation reference can be shuffled and mapped to the playout timeline. Once the structures are selected from a single or multiple time reference the structures can be shuffled in presentation time to satisfy a query. This is shown in Fig. 2.6.

Some techniques to achieve dynamic composition of video data have been accomplished. ConText [31] is a system for automatic temporal composition

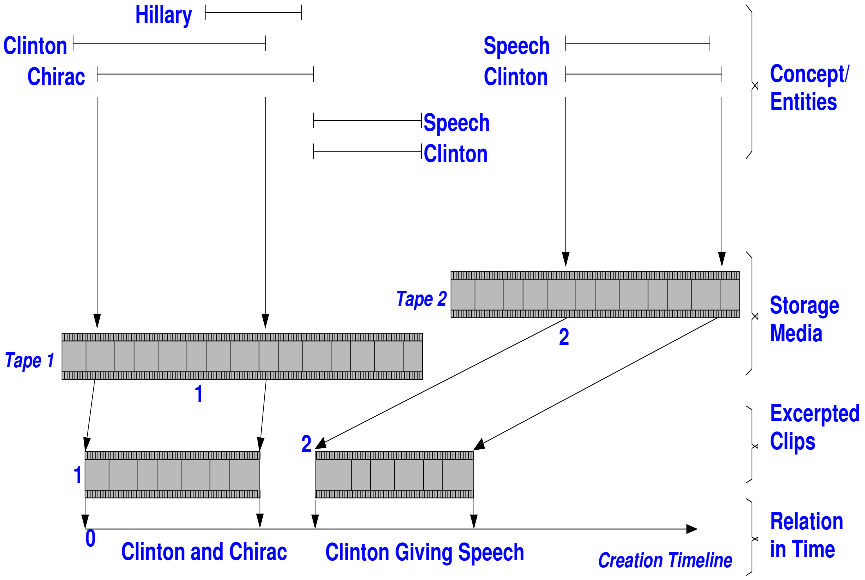


Figure 2.5: Spatio-Temporal Mapping Achieved by Structures from Multiple References

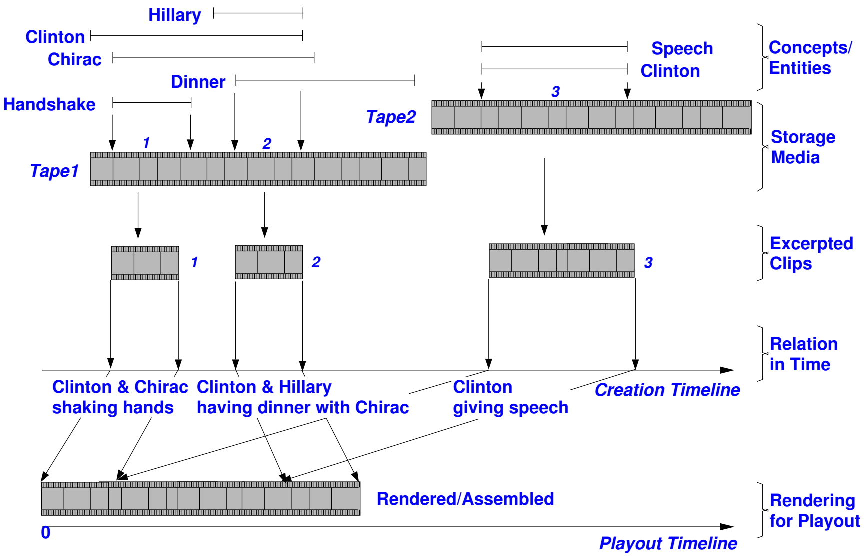


Figure 2.6: Spatio-Temporal Mapping Achieved by Shuffling the Structures

of a collection of video shots. It lets users navigate semi-randomly through a collection of documentary scenes associated with a limited range of content metadata describing *character*, *time*, *location*, and *theme*. The next scene shown to the user is determined based on a scoring of all available scenes. This scoring aims to obtain the preferred continuity and progression of detail in the presentation. This is made possible by establishing a present context consisting of metadata found in already-played shots or shots chosen by a user. Each metadata entry is associated with a relevance score. The theme, or storyline, is maintained by human intervention and is not completely automated.

ConText demonstrates how cognitive annotations of video material can be used to individualize a viewing session by creating an entirely new version through context-driven concatenation. This dynamic reconstruction can include video material made in a totally different context, thus performing a repurposing of the material. The temporal ordering in a composition is maintained by scoring the weights given to the keywords representing different types of information.

AUTEUR [64] is an application that is used to automatically generate humorous video sequences from arbitrary video material. The composition is based on the content describing the *characters*, *actions*, *moods*, and *locations*; as well as the information about the position of the camera with respect to a character, such as, close-up, medium, and long range shots. Content-based rules are used to compose shots.

Oomoto and Tanaka [68] use the concept of video object and the video model that consists of hierarchical composition of video based on content or descriptive information associated with the clips. Weiss et al. [92] propose composition based on video algebra. The video model used is similar to the previous work [68]. In addition they propose composition using algebraic operation, like union, concatenation, and intersection.

A number of systems have been proposed for delivery of news video data. These systems simply filter pre-composed news video. Agora [42], an application developed at Bell Labs, uses filtering of multi-channel broadcast news based on a user profile and closed-caption data. The Network Multimedia Information Services [28] system uses start and stop boundaries supplied with individual news items that can be browsed on the Web using an index. Shararay et al. at Bell Labs developed an application that maps the transcripts of the broadcast news with the associated video frames [82]. Transcripts can be browsed via the WWW browser. Brown et al. [20] also have a system that uses closed-caption text to filter information leading to the playout of associated clips.

The above composition techniques rely only on content for composition. Besides content, *structure* and *time* are also critical elements in a composition. A structure depicts cinematographic rules like establishing a starting scene, intermediate scenes, and a closing scene. Time maintains the temporal sequence of events, and is important for presenting information in the correct time-series. A news item, for instance, is a series of sub-events or a cause and effects chain, in which the time series must be maintained. In this dissertation we present composition techniques that take all the three features of content, structure and time into consideration.

Various types of information customization techniques have been proposed. These techniques use content, time, and cost specification and are discussed next.

## 2.3 Video Data Customization Techniques

In dynamic assembly of content it is possible to adapt the retrieved information to an individual's specification and a system's capabilities. Customization effects the retrieval, scheduling, and composition of a set of data [47]. Information for customization can be acquired either by implicit or explicit techniques. For explicit techniques [48] a user profile is acquired from the user directly. For implicit techniques [70, 84, 93] a user profile is acquired by observing the behavior of the user (e.g., information about a user's content preference and the order of the presentation). Techniques based on *user* profile [56], *society* or community profile to which a user belongs [54, 55, 70, 76, 83], and *economics* or cost and benefits of production and consumption [55] are used to achieve customization. These techniques are summarized below:

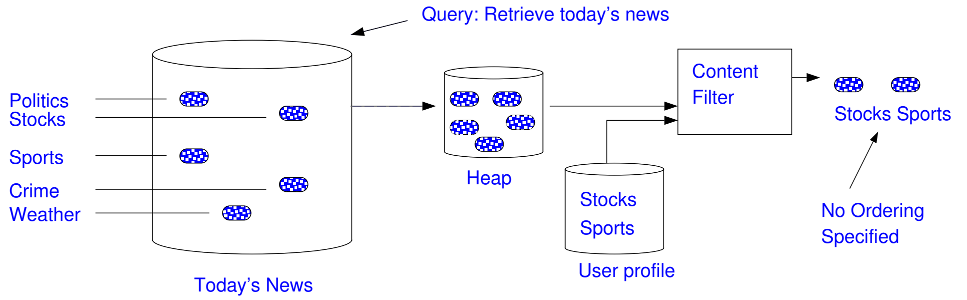


Figure 2.7: An Example of Content-Based Customization

**Content-based customization:** Only the preferred information is composed and rest of the information is dropped (content filtering) (Fig. 2.7).

**Cost-based customization:** Both the content provider and a user can specify cost parameters (e.g., quality of picture or price/unit time). Content provider is concerned with the profits while a user wants the best deal for minimum cost. Cost-based customization can depend on the value of intellectual property, network bandwidth, transmission time, data resolution required, or the age of the content (e.g., latest information is priced high).

**Structure-based customization:** We define this type of customization as filtering based on structural unit type (e.g., field shots). Fig. 2.8 illustrates an example of a structure-based customization, where only headlines are retained in the final composition.

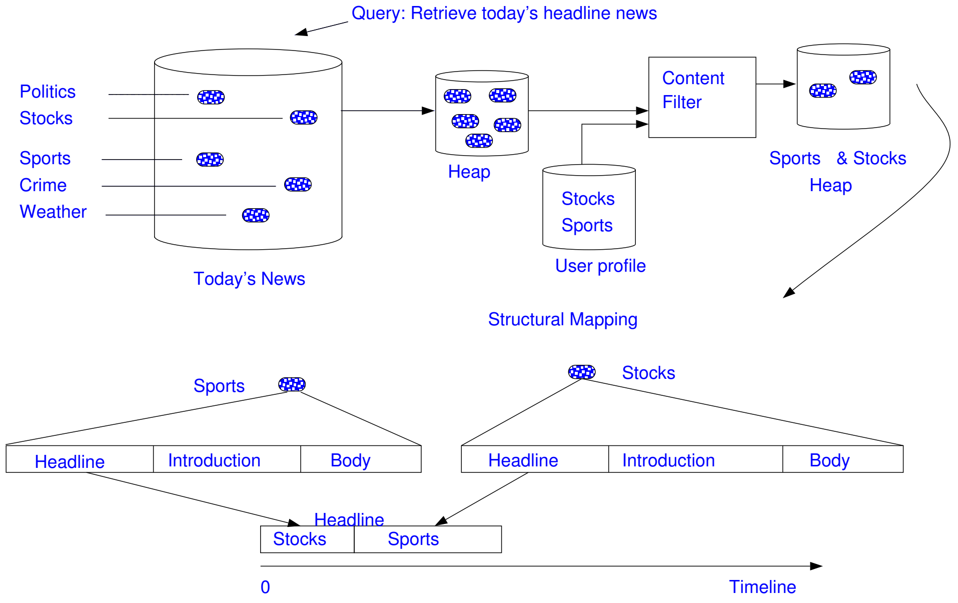


Figure 2.8: An Example of Structure-Based Customization

**Time-based customization:** There are two types of time-based customization: customization based on playout order and duration.

**Temporal order:** Customization is achieved by specification of the relative position of segments on a timeline as shown in Fig. 2.9. Depending upon the specifications, selected clips are mapped on a timeline for the playout. Initially, clips that match user profile are retained while the rest are filtered/dropped. Next, based on the temporal preference (i.e., stocks before sports), segments are mapped on the timeline for the playout.

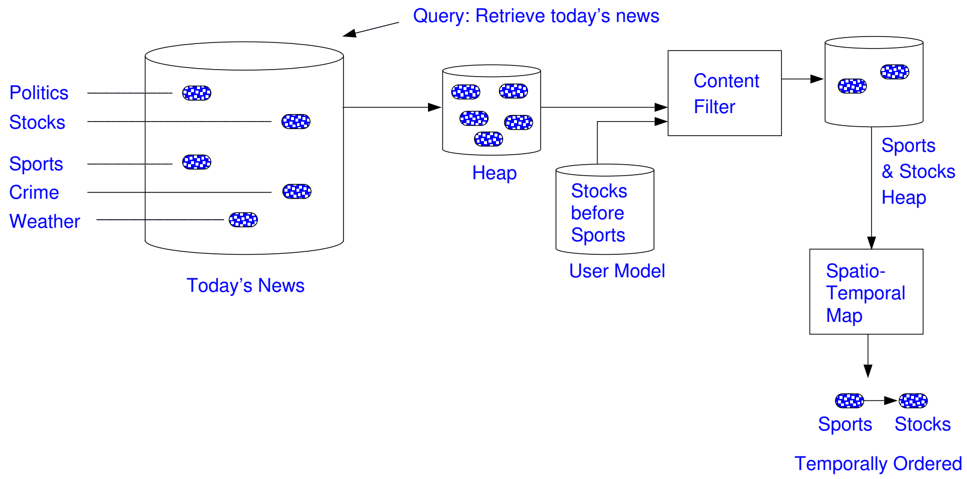


Figure 2.9: An Example of Time-Based Customization

**Time duration:** Customization is achieved by specification of the playout duration (e.g., the query "re-cap today's news for two minutes"). If the playout duration of the available data is more than the requested duration, some data need to be dropped. Currently the customization is achieved by imposing rules on content or information contained in video clips. For example, proof of a theorem cannot be presented before the problem statement. Ozsoyoglu et al. [72] impose content-based rules/constraints to drop or include the segments in a composition. A shortcoming of the content-based rules approach is the requirement of a set of rules for each and every scenario. The number of rules increases with the number of scenarios. It will not be possible to establish content-based rules when there are real-time requirements between acquisition and delivery of the incoming data.

The limitations of dropping data based on constraints imposed on content can be overcome by imposing rules based on the structure of an application domain. The advantage of this technique is that we require only a single set of rules that do not change (e.g., in news you cannot present details of a story without an introduction; or details can be dropped but not the introduction).

Further, Ozsoyoglu et al. [72] present algorithms for composing a multimedia (text, graphics, video, etc.) presentation in a specified duration. Depending on the content-based rules, first the multimedia data are composed in sub-presentations and then the proposed algorithms based on the requirement of the maximum number of windows to be open at any time and the total duration of the presentation are used to compose the multimedia presentation. This work has not been specifically targeted to the composition of video data and the authors do not discuss the quality of a resulting multimedia presentation.

Smith and Kanade [87] use a skimming technique to reduce the playout time of a composition for browsing. They identify significant audio and video information to create a synopsis. This work is more like creating a table-of-contents rather than a cohesive composition. Significant audio and video segments do not possess complete information, rather, indication of the available information.

Kamahara et al. [45] propose automatic program composition for a news-on-demand system. They present techniques for recomposition of data under the playout time constraints. They break a broadcast composition into unit data, where a unit data corresponds to data between two successive shot boundaries. Depending on the time specification, each unit is sequentially played out and stopped when time runs out, thus stopping at an arbitrary point in a composition. Therefore, this technique does not provide a cohesive composition.

In a composition, even if segments are dropped to adjust the playout time, the the resulting composition should still be a complete and cohesive composition. Furthermore, since segments consist of concepts from multiple perspectives, the time-constrained compositions should maintain the ability to present as much information from different perspectives as possible. In addition, the time-constrained composition should be able to cover as much creation period as possible. Therefore, we require composition and customization techniques that take these features into account. In addition, we require means to evaluate the quality of resulting compositions to compare them with manually composed video. These issues are discussed in detail in subsequent chapters.

## 2.4 Summary

In this chapter we have reviewed several existing metrics and techniques that are used for evaluation, information retrieval, and presentation of a video composition. In each case, we have highlighted the limitations of these metrics and techniques, and have identified specific areas of deficiency in evaluation and composition of a video piece. A general conclusion of our survey is that while many of the existing metrics can be used to evaluate the performance (recall & precision) of a retrieval system, they are not sufficient for evaluating the quality of video compositions. Similarly, we find that the existing composition techniques rely only on content for composition, and do not consider structure and time which are the other critical elements in a video composition.

Our review of the existing metrics and techniques for video composition has identified new areas for innovation and has highlighted specific areas for improvement. In particular, we find that the existing metrics that we are aware of offer little in terms of evaluation of a video composition. Existing metrics are useful only for retrieval of information, but do not consider various features like theme, temporal continuity, and structure in digital video production systems. Therefore, there is a need for development of a new set of metrics for evaluating the performance of a DVPS.

Our review shows that the existing video composition techniques are content-based only. In other words, while these techniques compose a video piece by finding the similarity among the concepts associated with video segments, they do not incorporate creation time and structure features. Creation time plays a critical role in maintaining temporal integrity of a video piece. For example, it is important that in a newscast the facts are presented in the correct chronological order since a news item can last over a time-period. Similarly, since not all segments of a video piece are alike, the overall structure of the composition is affected by the placements of these segments in the composition. For example, some segments are good candidates for staring a narrative while others better describe the event associated with the narrative. By not considering creation time and structure features of a video composition, existing techniques produce an inferior video piece. Therefore, there is a need for improvement in these aspects of video composition techniques.

In subsequent chapters we propose a set of new metrics for evaluation of a video composition. In addition, we also propose improved composition and customization techniques for digital video production systems.

# Chapter 3

# Metrics for Evaluation of a Composition

## Synopsis

In this chapter, we discuss the features that represent a video composition. These features include information contained in a composition, information flow in a composition, temporal ordering of content in a composition, structure of information, and creation time period of content in a composition. Based on this feature set, we formulate a set of metrics that are used to quantify a dynamic video composition. We demonstrate the use of these metrics with help of examples. The proposed metrics are used to evaluate the quality of manually composed news broadcast videos, and the results establish the baseline reference values for automated news video compositions.

## 3.1 Introduction

To convey a story using the video medium requires a succession of video segments corresponding to concepts of its narrative. The narrative also has a main concept, or focus, called the story center. Therefore, a story is achieved by the composition of a succession of video segments mapping concepts or *threads* that include the story center and multiple related concepts. To quantify the character of the video segments, we identify some fundamental attributes, or the feature set, of video narratives.

The first attribute is *temporal continuity*, which characterizes the sequencing of segments in time. A video composition is created by composing information about a story or story center; it shows changes as the story develops and progresses. In other words, a composition is a chain of cause and effects. Therefore, the position of a particular cause or effect in a composition is very important. The information needs to be presented along a timeline, for example, a scoring time series in a game. The quality of the composition is also effected by the position of a segment on a timeline. We cannot *transpose* older facts to a position in future without first introducing a change in context.

The next attribute is *thematic continuity*, or the smooth flow of conveyed information between consecutive segments. In a composition different views or perspectives are present about a story or story center. For example, multiple views of an event are presented in a news item (e.g., field shots and interviews). Therefore, there are different sets of segments that present

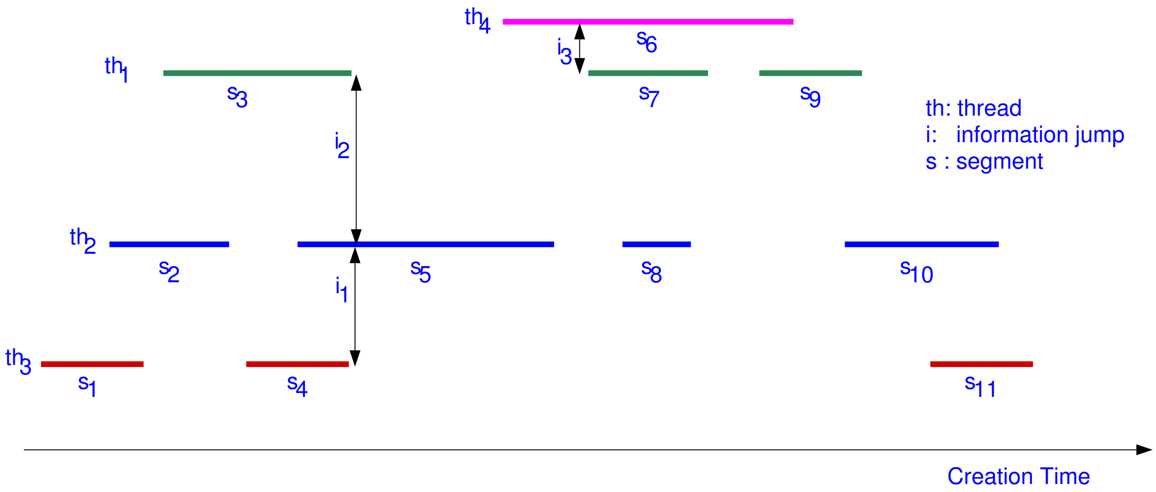


Figure 3.1: Schematic of Threads in a Composition

domain-relevant information but by different vehicles. The sets possessing temporally-ordered segments are called *threads*, where each thread contains information from a different perspective about an event. Fig. 3.1 graphically illustrates this concept of threads. Each thread induces a thematic jump, or shift in the theme of the story. Hence, the segments associated with the threads must be ordered to maintain overall continuity in theme throughout a composition.

Another attribute is *period span coverage*. The lifespan of an event can vary from a single day to many years. A composition can encompass this entire period or a subset of this period. We describe and quantify this coverage as period span coverage. We also consider the continuity of types of assembled components of the composition. For example, a news item has structure that consists of an introduction, a body, and an end. A composed video piece should conform to such a domain-dependent structure. This attribute is described as *structural continuity*.

*Content progression* in a composition also plays an important role. A consumer must be able to assimilate the contents of each segment within its duration, yet should not be presented with unnecessary content. This must be balanced with the exclusion of *information* that can be lost when segments are shortened or dropped from a composition. Here we define information as the sum of the concepts encompassed in the composition.

The feature set for characterizing video compositions consists of information, thematic continuity, temporal continuity, structural continuity, period span coverage, and content progression. Next, we formulate techniques to quantify these attributes. The symbols used in this chapter and later chapters are summarized in Tables 3.1, 3.2, 3.3, and 3.7.

## 3.2 Metrics

We propose a metric for each attribute in the feature set. The formulation of the metrics assumes the existence of a *candidate set* $S_a$ of segments for a composition. That is, the candidate set of segments $S_a$ from the universe of available video segments, $S$, satisfies a particular selection criterion. Ultimately, the candidate set yields a *composition set* $S_c$ which, when ordered, comprises the final video composition. Intuitively, $S_c \subseteq S_a \subseteq S$.

To support characterization of the video segments, we define a tuple $< b, \vec{W}, d >$, where $b$ is the creation time and date of the segment, $d$ is the

Table 3.1: Symbols Used to Define Segments and Sets

| **Symbol** | **Description** |
|---|---|
| $s$ | Segment |
| $S$ | Universe of video segments |
| $N$ | Size of the segment universe |
| $b$ | Creation time and date of segment $s$ |
| $C$ | Universe of concepts |
| $d$ | Playout duration of a segment $s$ |
| $S_a$ | Candidate set |
| $N_a$ | Size of the candidate set |
| $S_c$ | Composition set |
| $N_c$ | Size of the composition set |
| $S_c^k$ | $k$th set of composition segments (multiple compositions) |

playout duration of the segment, and $\vec{W}$ is an ordered set of concepts for the segment with respect to the universe of concepts, $C$, contained in $S$.

Sets $S_a$ and $S_c$ are refinements on $S$ that lead to the composition. These refinements are performed in practice by database queries performing similarity matching between user-input interest criteria and the set of concepts associated with each segment in $S$. The concepts associated with each segment are established during annotation (upon inclusion in $S$).

Table 3.2: Symbols Used to Define Concept Vectors

| **Symbol** | **Description** |
|---|---|
| $\vec{W}$ | Concept weight vector for a segment |
| $w_i$ | Weight associated with concept $c_i$ |
| $\bar{w}_i^a$ | Average weight associated with a concept $c_i$ for a *candidate* set |
| $\bar{w}_i^c$ | Average weight associated with a concept $c_i$ for a *composition* set |
| $\vec{C}_a$ | Centroid vector for a *candidate* set |
| $\vec{C}_c$ | Centroid vector for a *composed* set |

To simplify the mathematics, we make two assumptions about $S$. First, we assume that both $|S|$ and $|C|$ are constant during evaluation. Second, we assume that $S$ has a chronological order of creation times. This property can be achieved by the mapping $M$ from the set of natural numbers to segments in $S$, where $M$ is one of the permutations of the set of natural numbers:

$$\exists M : M \subset \mathcal{S}_N : (\forall i : 1 \leq i < N : b_{M(i)} \leq b_{M(i+1)}), \tag{3.1}$$

where $N = |S|$, $\mathcal{S}_N$ is a symmetric group of permutations of degree $N$, and

$M$ is as defined above. The relation $M$ permits segments to be chronologically ordered by creation time independently from subscript values. For the remainder of the paper, our use of the term "consecutive segments" implies this property of adjacency in creation times.

The metrics are described below:

Table 3.3: Symbols Used to Define Metrics

| **Symbol** | **Description** |
|---|---|
| $In$ | Information |
| $e_{tc}$ | Temporal continuity |
| $e_{thc}$ | Thematic continuity |
| $e_{cp}$ | Content progression |
| $e_{sc}$ | Structural continuity |
| $e_{ps}$ | Period span |
| $\beta$ | Forward jump weight for temporal continuity |
| $\delta$ | Forward jump tolerance |
| $\lambda$ | Dissimilarity threshold |
| $\tau$ | Similarity threshold |
| $\rho$ | Fast change threshold |
| $\varrho$ | Slow change threshold |
| $D_t$ | Target temporal span |
| $D_a$ | Achieved temporal span |

### Information

This metric measures the amount of information, or the sum of the concepts represented in a composition (these associated segments comprise the composition set ($S_c$)), as compared to the information available in the candidate set ($S_a$). We calculate the amount of information in a composition as follows.

We define $\vec{W} = [w_1,\ w_2,\ w_3, ..., w_{|C|}]$ as the concept weight vector characterizing the weight of each concept in the concept universe associated with a segment $s$. (These weights are defined at the time that $s$ enters $S$ through manual or automatic techniques.). A *centroid* vector is defined as $\vec{C} = [\bar{w}_1, \bar{w}_2, \bar{w}_3, ..., \bar{w}_{|C|}]$ where each $\bar{w}$ represents the average weight of a concept from the represented segments in the set. Subscripts $a$ and $c$ are used to describe candidate or composition sets in this notation. Therefore,

$$\bar{w}_i^a \ = \ \frac{1}{N_a} \sum_{\forall s \in S_a} w_i$$

represents the average weight of concept $w_i$ for elements in the candidate set $S_a$ that form the centroid vector $\vec{C_a}$. The centroid vector for the composition set ($\vec{C_c}$) is similarly defined on $S_c$.

To evaluate the information metric, we measure the similarity of information between $\vec{C_a}$ and $\vec{C_c}$ using the cosine similarity metric proposed by Salton [79]. This technique measures the distance between the two vectors in the concept space of dimension $n$:

$$\text{cosine}(\vec{A}, \vec{B}) = \frac{\sum_{k=1}^{n}(a_k \times b_k)}{\sqrt{\sum_{k=1}^{n}(a_k)^2 \times \sum_{k=1}^{n}(b_k)^2}}$$

Applying this technique, the information metric, $In$, is defined as:

$$In = \frac{\text{cosine}(\vec{C}_a, \vec{C}_c) \times N_c}{N_a}.$$

We observe that the weight of a concept central to a storyline does not vary appreciably in a candidate set and the cosine value by itself is not sensitive to the concepts occurring less frequently in a composition. Therefore, we scale the cosine value with the factor $\frac{N_c}{N_a}$. If the information in the two vectors is the same, then $In = 1$, otherwise, $In < 1$.

This metric can be evaluated using, for example, the data of Fig. 3.2. Consider the set of video segments and their concept vectors with binary weights as shown in the figure. If all segments are incorporated in the composition then the centroid vectors, of the candidate set and composition are:

$$\vec{C}_a = \vec{C}_c = [0.72\ 0.72\ 0.54\ 0.63\ 0.45\ 0.18\ 0.54\ 0.09\ 0.27\ 0.27\ 0.18\ 0.18\ 0.09\ 0.18\ 0.27\ 0.18\ 0.27\ 0.27]$$

The information value ($In$) of the two vectors is equal to 1. Suppose segments $s_5$ to $s_8$ are not in the composition, then the centroid vector for the candidate set and composition are:

$$\vec{C}_a = [0.72\ 0.72\ 0.54\ 0.63\ 0.45\ 0.18\ 0.54\ 0.09\ 0.27\ 0.27\ 0.18\ 0.18\ 0.09\ 0.18\ 0.27\ 0.18\ 0.27\ 0.27]$$

$$\vec{C}_c = [0.85\ 0.57\ 0.57\ 0.57\ 0.57\ 0.28\ 0.57\ 0.00\ 0.14\ 0.14\ 0.28\ 0.28\ 0.14\ 0.28\ 0.28\ 0.14\ 0.28\ 0.14]$$

The information value, $In$ is now equal to 0.61, and there is a 39% reduction in the value of $In$.

|  | $s_1$ | $s_2$ | $s_3$ | $s_4$ | $s_5$ | $s_6$ | $s_7$ | $s_8$ | $s_9$ | $s_{10}$ | $s_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| India | 1 | 1 | 1 |  |  | 1 |  | 1 | 1 | 1 | 1 |
| Nuclear | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |  |  |  |
| Bomb | 1 | 1 | 1 | 1 | 1 |  |  | 1 |  |  |  |
| Test | 1 | 1 | 1 | 1 | 1 | 1 | 1 |  |  |  |  |
| Fission | 1 | 1 | 1 | 1 | 1 |  |  |  |  |  |  |
| Underground | 1 |  |  | 1 |  |  |  |  |  |  |  |
| Pakistan |  |  |  | 1 | 1 |  | 1 |  | 1 | 1 | 1 |
| China |  |  |  |  |  |  |  | 1 |  |  |  |
| CTBT |  |  |  |  |  | 1 | 1 |  |  |  | 1 |
| NTP |  |  |  |  |  | 1 | 1 |  |  |  | 1 |
| Kashmir |  |  |  |  |  |  |  |  | 1 | 1 |  |
| Dialogue |  |  |  |  |  |  |  |  | 1 | 1 |  |
| Bilateral |  |  |  |  |  |  |  |  | 1 |  |  |
| UN |  |  | 1 |  |  |  |  |  |  | 1 |  |
| USA |  | 1 |  |  | 1 |  |  |  |  |  | 1 |
| Sanctions |  | 1 |  |  | 1 |  |  |  |  |  |  |
| Condemned |  | 1 | 1 |  | 1 |  |  |  |  |  |  |
| Amend |  |  |  |  |  | 1 | 1 |  |  |  | 1 |

Figure 3.2: Example Concept Vector Weights

### Temporal Continuity

Temporal continuity, $e_{tc}$, is quantified as follows: let $N_c$ represent the number of segments placed on the creation timeline, and the distances between segments be measured in time. Let large forward jumps in time (if such data exist) be less damaging to temporal continuity than reverse jumps, and let forward jumps be weighted by $0 \leq \beta \leq 1$. We define *good* temporal continuity to mean that all cause-effects in a story follow an increasing time series.

**Temporal Continuity:**

$$\begin{array}{lll} 0 \leq b_{i+1} - b_i \ \leq \delta & \Rightarrow & e^i_{tc} = 1 \\ b_{i+1} - b_i \ > \delta & \Rightarrow & e^i_{tc} = 1 - \beta((b_{i+1} - b_i) - \delta)/D_t \\ b_{i+1} - b_i \ < 0 & \Rightarrow & e^i_{tc} = 1 - (b_i - b_{i+1})/D_t \end{array}$$

Here, $\delta$ is the duration that can be tolerated in a forward jump and $D_t$ is the target temporal span of the data. The mean temporal continuity of the segments on the timeline is $\frac{1}{N_c-1}\sum_{i=1}^{N_c-1} e^i_{tc}$.

In Fig. 3.3 the behavior of thematic continuity is shown. The creation time and date $b_i$ of is always taken as 0 hours and creation time and date $b_{i+1}$ is changed in the increments of both 24 hours and -24 hours. $D_t$ is taken as 960 hours, $\delta$ is taken as 24 hours, and the weight $\beta$ is taken as 0.6. As seen from the figure the temporal continuity with reverse jumps is penalized more then the forward jumps.

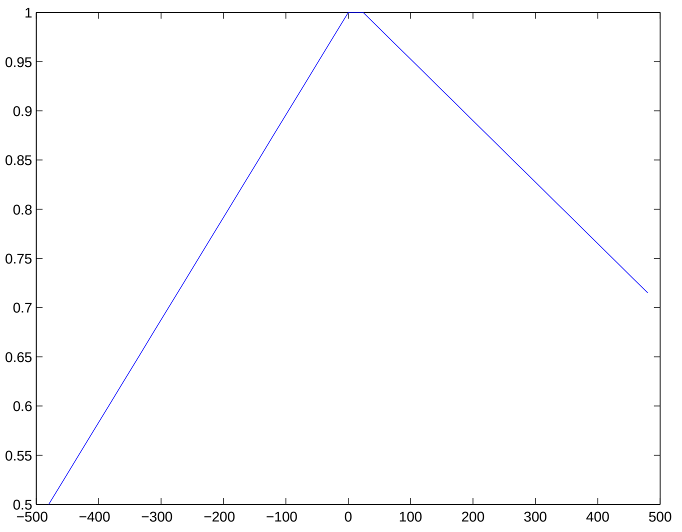


Figure 3.3: Schematic Behavior of the Temporal Continuity Metric

Consider the creation time and date of the segments of Fig. 3.2 as shown in Table 3.4. Assume that the tolerated jump duration, $\delta$, is 24 hours and weight $\beta$ is 0.6 and consider the playout sequence $[s_1\,, s_2\,, s_3\,, s_4\,, s_5\,, s_6\,, s_7\,, s_8\,, s_9\,, s_{10}\,, s_{11}]$.

All jumps in this example are forward in time and less than $\delta$ in duration with the exception of the jump between $s_5$ and $s_6$. However, since there are no data corresponding to this jump window, there is no penalty. Gaps in data are usually due to the news item being off-air for long periods due to lack of new developments. As the $e_{tc}$ for all consecutive pairs of segments is 1, the mean temporal continuity is also equal to 1. Consider the sequence $[s_1\,, s_2\,, s_4\,, s_5\,, s_6\,, s_7\,, s_8\,, s_{10}\,, s_{11}\,, s_9]$. The $e_{tc}$ for this sequence is shown in Table 3.4 and the mean temporal continuity is 0.97.

Table 3.4: Creation Time, Date, and Temporal Continuity

| **Segment** | **Time** | **Date** |
|---|---|---|
| $s_1$ | 08:00:00 | 01/12/98 |
| $s_2$ | 06:30:00 | 01/13/98 |
| $s_3$ | 22:00:00 | 01/13/98 |
| $s_4$ | 10:00:00 | 01/14/98 |
| $s_5$ | 08:00:00 | 01/15/98 |
| $s_6$ | 20:00:00 | 01/16/98 |
| $s_7$ | 14:00:00 | 01/17/98 |
| $s_8$ | 12:00:00 | 01/18/98 |
| $s_9$ | 22:00:00 | 01/18/98 |
| $s_{10}$ | 14:00:00 | 01/19/98 |
| $s_{11}$ | 08:00:00 | 01/20/98 |

| **Segments** | $e_{tc}$ |
|---|---:|
| $s_1 - s_2$ | 1 |
| $s_2 - s_4$ | $1 - 0.6(1650 - 1440)/7 \times 24 \times 60 = 0.98$ |
| $s_4 - s_5$ | 1 |
| $s_5 - s_6$ | 1 |
| $s_6 - s_7$ | 1 |
| $s_7 - s_8$ | 1 |
| $s_8 - s_{10}$ | $1 - 0.6(1560 - 1440)/7 \times 24 \times 60 = 0.99$ |
| $s_{10} - s_{11}$ | 1 |
| $s_{11} - s_9$ | $1 - (2160)/7 \times 24 \times 60 = 0.78$ |

### Thematic Continuity

This metric, $e_{thc}$, quantifies the progression of a storyline or a theme in a composition. We establish a similarity threshold $\tau$, and if the similarity measure between the two consecutive segments is more than $\tau$, the two segments are considered very similar and progression of the theme is static. We also establish a dissimilarity threshold $\lambda$, below which segments are considered disjoint.

**Thematic Continuity:**

$$
\begin{array}{lll}
\lambda \ \leq \ \text{cosine}(\vec{W}_i, \vec{W}_{i+1}) \ \leq \tau & \Rightarrow & e^i_{thc} = 1 \\
\text{cosine}(\vec{W}_i, \vec{W}_{i+1}) \ > \tau & \Rightarrow & e^i_{thc} = \frac{\tau}{\text{cosine}(\vec{W}_i, \vec{W}_{i+1})} \\
\text{cosine}(\vec{W}_i, \vec{W}_{i+1}) \ < \lambda & \Rightarrow & e^i_{thc} = \frac{\text{cosine}(\vec{W}_i, \vec{W}_{i+1})}{\lambda}
\end{array}
$$

The mean thematic continuity of a composition is $\frac{1}{N_c-1}\sum_{i=1}^{N_c-1} e^i_{thc}$. In the Fig. 3.4 the schematic behavior of thematic continuity with dissimilarity threshold $\lambda$ of 0.4 and similarity threshold $\tau$ of 0.7 is shown.

Consider a dissimilarity threshold $\lambda$ of 0.6, similarity threshold $\tau$ of 0.9, and a sequence of $[s_1\,, s_2\,, s_3\,, s_4\,, s_5\,, s_6\,, s_{10}\,, s_{11}]$. The thematic continuity of the composition is calculated in steps using the concept vectors of Fig. 3.2 and is shown in Table 3.5. The final result is a value of 0.76.

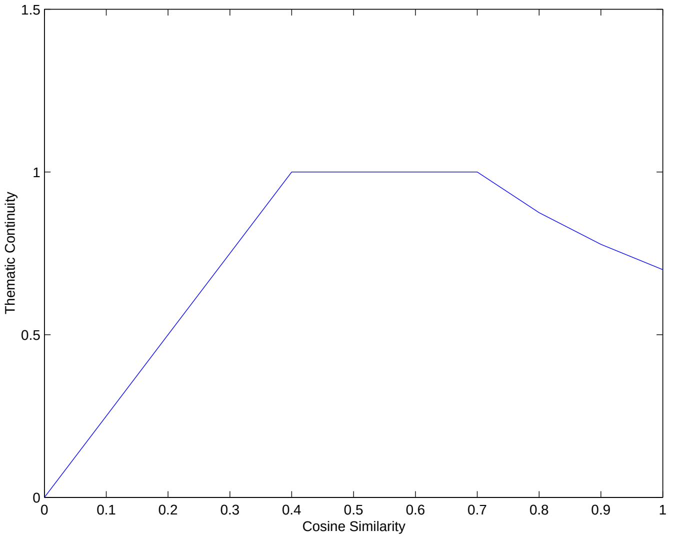


Figure 3.4: Schematic Behavior of the Thematic Continuity Metric

Table 3.5: Thematic Continuity

| **Segments** | **cosine** | $e_{thc}$ |
|---|---|---|
| $s_1 - s_2$ | $5/\sqrt{6 \times 8} = 0.72$ | 1 |
| $s_2 - s_3$ | $6/\sqrt{8 \times 7} = 0.80$ | 1 |
| $s_3 - s_4$ | $4/\sqrt{7 \times 6} = 0.61$ | 1 |
| $s_4 - s_5$ | $5/\sqrt{6 \times 8} = 0.72$ | 1 |
| $s_5 - s_6$ | $2/\sqrt{8 \times 6} = 0.28$ | $0.28/0.6 = 0.46$ |
| $s_6 - s_{10}$ | $1/\sqrt{6 \times 5} = 0.18$ | $0.18/0.6 = 0.3$ |
| $s_{10} - s_{11}$ | $2/\sqrt{5 \times 6} = 0.36$ | $0.36/0.6 = 0.6$ |

### Content Progression

This metric, $e_{cp}$, characterizes the rate at which concepts change within a composition. Changes that are too fast or too slow deteriorate the quality of a composition.

We consider content progression as being *fast* if, given that there are variations in the information contained in consecutive segments, the duration of playout of the consecutive segments is smaller than a fast-change threshold $\rho$. If the playout duration of a segment is greater than a slow-change threshold, $\varrho$, then a long time is consumed on discussing a certain aspect of an event and the content progression is considered *slow*. The content progression is measured as follows:

**Content Progression:**

$$\begin{array}{lll} \rho \ \leq \ d_i \ \leq \varrho & \Rightarrow & e^i_{cp} = 1 \\ d_i \ > \ \varrho & \Rightarrow & e^i_{cp} = \frac{\varrho}{d_i} \\ d_i \ < \ \rho & \Rightarrow & e^i_{cp} = \frac{d_i}{\rho} \end{array}$$

Here, $e_{cp}$ is defined as progression continuity and $d_i$ is the playout duration of segment $s_i$. The mean playout duration of the segments is $\frac{1}{N_c}\sum_{i=1}^{N_c} e^i_{cp}$.

The schematic behavior of the content progression metric is similar to the thematic continuity metrics (Fig. 3.5

Consider the playout durations of the segments shown in Table 3.6. Assume a fast-change threshold $\rho$ of 5 seconds and a slow-change threshold $\varrho$ of 150 seconds. The content progression of the sequence $[s_1\,, s_2\,, s_3\,, s_4\,, s_5\,, s_6\,, s_{10}\,, s_{11}]$

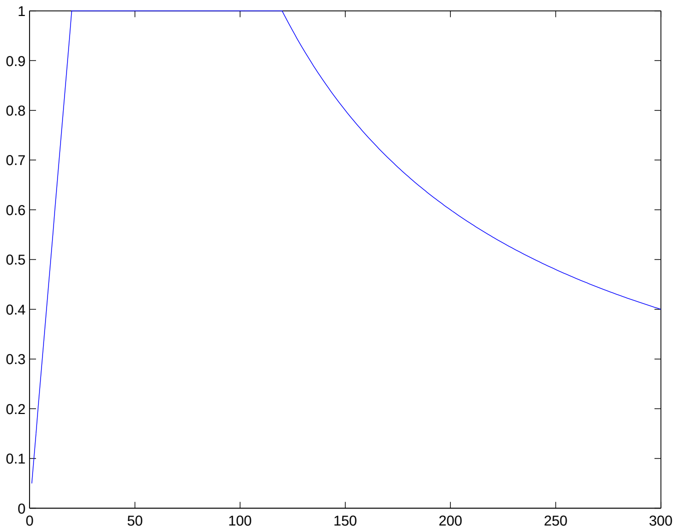


Figure 3.5: Schematic Behavior of the Content Progression Metric

is shown in Table 3.6. The mean content progression of the sequence evaluates to 0.81.

### Period Span Coverage

This metric quantifies the performance of a system for covering information from a complete period for which data are available and selected. Let $D_t$ be the target span requested for composition and $D_a$ be the span covered by segments in the data universe under the selection criterion. Period span coverage, $e_{ps}$, is defined as $\frac{D_a}{D_t}$.

Consider the segments summarized in Table 3.4. The complete span of the data in the table is from 12 Jan 1998 to 20 Jan 1998. For the sequence $\{s_1, s_2, s_4\}$, the span coverage of the composition is $2/8 = 0.25$.

Table 3.6: Playout Duration and Content Progression

| **Segment** | **Duration (s)** |
|---|---:|
| $s_1$ | 10 |
| $s_2$ | 15 |
| $s_3$ | 2 |
| $s_4$ | 3 |
| $s_5$ | 30 |
| $s_6$ | 60 |
| $s_7$ | 12 |
| $s_8$ | 4 |
| $s_9$ | 5 |
| $s_{10}$ | 120 |
| $s_{11}$ | 300 |

| **Segment** | $e_{cp}$ |
|---|---:|
| $s_1$ | 1 |
| $s_2$ | 1 |
| $s_3$ | $2/5 = 0.4$ |
| $s_4$ | $3/5 = 0.6$ |
| $s_5$ | 1 |
| $s_6$ | 1 |
| $s_{10}$ | 1 |
| $s_{11}$ | $150/300 = 0.5$ |

Table 3.7: Symbols Used to Define News Video Segment Types

| **Symbol** | **Description** |
|---|---|
| $S_h$ | Set of Headline-type segments |
| $S_{in}$ | Set of Introduction-type segments |
| $S_b$ | Set of news body-type segments |
| $S_e$ | Set of Enclose-type segments |

### Structural Continuity

The structural continuity metric is defined with respect to an established domain-specific structure, and quantifies deviation. Below, we describe a structural continuity metric for broadcast news video. The evaluation is binary; degrees of discontinuity can be defined but are not considered here.

**Structural Continuity for News Items:**

$\{s_h\} = C \Rightarrow e_{sc} = 1$ Only a headline can be present in a composition $C$.

$\{s_h, s_{in}\} = C \Rightarrow e_{sc} = 1$ Only a headline and an introduction can be present in a composition $C$.

$\{s_{in}\} = C \Rightarrow e_{sc} = 1$ Only an introduction can be present in a composition.

$\{s_h, s_{in}, \{s_b^1, s_b^2, ...\}\} = C \Rightarrow e_{sc} = 1$ Only a headline, an introduction, and segments belonging to the body can be present in a composition.

$\{s_{in}, \{s_b^1, s_b^2, ...\}\} = C \Rightarrow e_{sc} = 1$ Only a headline and segments belonging to the body can be present.

$\{s_{in}, \{s_b^1, s_b^2, ...\}, s_e\} = C \Rightarrow e_{sc} = 1$ Only an introduction, segments belonging to the body, and an enclose can be present.

$\{s_h, s_{in}, \{s_b^1, s_b^2, ...\}, s_e\} = C \Rightarrow e_{sc} = 1$ All the segment types are present.

$\{\text{All other combinations}\} \Rightarrow e_{sc} = 0$

With the definition of these metrics for evaluation of video composition, we are prepared to establish reference values for manually-produced video in a specific domain. We use these reference values to evaluate the performance of our automatic composition techniques.

## 3.3 Analysis of a Broadcast News Composition

Broadcast news video production presents us with well defined domain-specific structures on which to apply our techniques. It is also readily available in adequate quantities. In the following we collect data from three broadcast sources and evaluate the quality of the broadcast news using our metrics. Details of data collection, analysis and results of the evaluation are described below.

**News Video Data Collection**

Broadcast news video data were acquired from CNN, NBC, and ABC over a period of 40 days from 20 January 1998 to 28 February 1998. During this period we recorded the 9:00 AM and 8:00 PM CNN (national) broadcasts (CNN1 and CNN2), the 6:30 PM NBC (national) broadcast, and the 12:00 PM ABC (local) broadcast.

Data were initially recorded in analog, VHS/NTSC, format and considerable effort was required to translate the data into a state suitable for resolving queries to yield candidate sets and composable segments. The analog video streams were first digitized into MPEG-1 format and then content and structural information/metadata were extracted. Segments were annotated based on the types of components within each news item. Content information such as conceptual and tangible entities [6] (e.g., people, locations, cause and effects, and events) were annotated to support the generation of concept vectors. Based on this data set we applied our metrics.

### Thematic Continuity and Content Progression

The thematic continuity ($e_{tc}$) was evaluated with a dissimilarity threshold $\lambda = 0.6$ and a similarity threshold $\tau = 0.9$. The content progression ($e_{cp}$) was measured with a fast-change threshold $\rho = 8$ seconds and slow-change threshold $\varrho = 100$ seconds. The results are summarized in Table 3.8.

The measurements show a thematic continuity that varies between 0.50 and 1.0. The low values indicate rough transitions between consecutive video segments. This is also apparent from a visual inspection of the corresponding segments where there are abrupt jumps in information level between threads of the news items. The content progression varies between 0.81 and 1.0. On average the playout duration of a segment is within the lower and upper limits set for measurement and there is a gradual change in content throughout the composition.

### Temporal Continuity

For measuring temporal continuity we assume that the creation time of a segment is the time when it is first shown in a composition. As mentioned before, segments transposed in time or segments with significant inter-segment tem-

Table 3.8: Thematic Continuity and Content Progression Measurements

| **News Item** | **No. of Segs** | $e_{thc}$ | $e_{cp}$ | **News Item** | **No. of Segs** | $e_{thc}$ | $e_{cp}$ | **News Item** | **No. of Segs** | $e_{thc}$ | $e_{cp}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 9 | 0.80 | 0.91 | 34 | 5 | 0.95 | 0.95 | 67 | 2 | 1.0 | 0.93 |
| 2 | 11 | 0.93 | 0.93 | 35 | 6 | 1.0 | 1.0 | 68 | 3 | 0.74 | 1.0 |
| 3 | 8 | 0.89 | 0.97 | 36 | 4 | 1.0 | 0.93 | 69 | 5 | 0.98 | 1.0 |
| 4 | 7 | 0.89 | 1.0 | 37 | 6 | 1.0 | 1.0 | 70 | 14 | 1.0 | 0.85 |
| 5 | 5 | 0.68 | 0.87 | 38 | 3 | 0.95 | 1.0 | 71 | 12 | 0.99 | 0.97 |
| 6 | 2 | 1.0 | 0.88 | 39 | 10 | 1.0 | 0.95 | 72 | 8 | 1.0 | 1.0 |
| 7 | 10 | 0.94 | 0.93 | 40 | 7 | 1.0 | 0.94 | 73 | 3 | 0.92 | 0.87 |
| 8 | 8 | 1.0 | 1.0 | 41 | 3 | 1.0 | 1.0 | 74 | 7 | 0.98 | 0.96 |
| 9 | 6 | 0.99 | 1.0 | 42 | 5 | 1.0 | 0.95 | 75 | 10 | 1.0 | 0.83 |
| 10 | 7 | 0.73 | 0.96 | 43 | 5 | 0.98 | 1.0 | 76 | 2 | 1.0 | 0.93 |
| 11 | 5 | 0.98 | 0.92 | 44 | 7 | 1.0 | 1.0 | 77 | 6 | 0.98 | 1.0 |
| 12 | 7 | 0.98 | 1.0 | 45 | 11 | 0.98 | 1.0 | 78 | 4 | 1.0 | 1.0 |
| 13 | 5 | 1.0 | 0.94 | 46 | 9 | 1.0 | 1.0 | 79 | 4 | 1.0 | 0.87 |
| 14 | 5 | 0.92 | 0.92 | 47 | 11 | 1.0 | 0.95 | 80 | 2 | 0.98 | 0.93 |
| 15 | 3 | 0.85 | 0.87 | 48 | 4 | 0.94 | 1.0 | 81 | 7 | 1.0 | 1.0 |
| 16 | 6 | 0.94 | 1.0 | 49 | 12 | 1.0 | 0.93 | 82 | 7 | 1.0 | 0.96 |
| 17 | 3 | 0.52 | 1.0 | 50 | 7 | 0.98 | 0.98 | 83 | 6 | 1.0 | 1.0 |
| 18 | 10 | 0.98 | 0.93 | 51 | 2 | 0.97 | 0.81 | 84 | 15 | 1.0 | 0.72 |
| 19 | 6 | 0.99 | 0.95 | 52 | 4 | 1.0 | 0.87 | 85 | 10 | 1.0 | 0.95 |
| 20 | 4 | 1.0 | 0.90 | 53 | 2 | 0.94 | 1.0 | 86 | 5 | 1.0 | 1.0 |
| 21 | 4 | 0.98 | 0.86 | 54 | 4 | 0.97 | 1.0 | 87 | 9 | 1.0 | 0.88 |
| 22 | 7 | 1.0 | 0.96 | 55 | 4 | 1.0 | 0.96 | 88 | 4 | 1.0 | 0.78 |
| 23 | 5 | 1.0 | 1.0 | 56 | 5 | .0 | 1.0 | 89 | 8 | 1.0 | 0.87 |
| 24 | 6 | 0.95 | 0.88 | 57 | 10 | 1.0 | 0.98 | 90 | 4 | 1.0 | 0.84 |
| 25 | 2 | 1.0 | 0.81 | 58 | 3 | 1.0 | 1.0 | 91 | 7 | 0.97 | 0.94 |
| 26 | 4 | 1.0 | 1.0 | 59 | 8 | 0.99 | 1.0 | 92 | 8 | 0.91 | 0.89 |
| 27 | 8 | 0.97 | 1.0 | 60 | 6 | 0.98 | 1.0 | 93 | 5 | 0.96 | 0.87 |
| 28 | 8 | 0.97 | 0.92 | 61 | 5 | 0.89 | 1.0 | 94 | 2 | 1.0 | 0.87 |
| 29 | 8 | 0.90 | 0.89 | 62 | 2 | 0.50 | 1.0 | 95 | 4 | 1.0 | 0.93 |
| 30 | 8 | 0.97 | 0.98 | 63 | 4 | 1.0 | 0.93 | 96 | 2 | 1.0 | 1.0 |
| 31 | 6 | 1.0 | 0.92 | 64 | 2 | 1.0 | 1.0 | 97 | 2 | 0.66 | 1.0 |
| 32 | 6 | 1.0 | 0.95 | 65 | 6 | 1.0 | 1.0 | 98 | 2 | 0.82 | 0.81 |
| 33 | 4 | 0.86 | 0.93 | 66 | 4 | 1.0 | 1.0 | 99 | 6 | 0.96 | 1.0 |

poral spans will yield low temporal continuity. To study this characteristic we isolated a single news event on the topic of the United Nations and Iraq Standoff.

We define inter-transposition duration as the period between segment repetition. The distribution of inter-transposition durations frequencies of the news items from CNN is shown in Fig. 3.6.

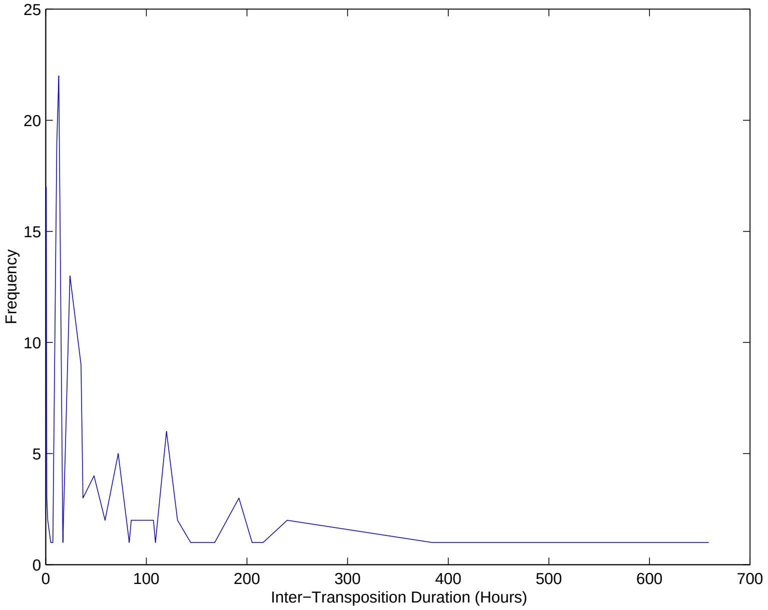


Figure 3.6: Inter-Transposition Durations for News Items from CNN1 and CNN2

The minimum inter-transposition time found in the result data set (Table 3.9 and Fig. 3.6) is less than one hour (i.e., the same segment is repeated in a single broadcast). On average, the interval between segment repetition (of 96 segments) from a single source (CNN) is 59 hours. The maximum inter-transposition time is 659 hours. Note that for this analysis we ignored segments that might have occurred prior to our observation period. Table 3.9 shows additional data characterizing the other news sources.

Table 3.9: Segment Inter-transposition Repetition History

| **Source** | **Number of Segments Repeated** | **Maximum Times Repeated** | **Average Inter-Transposition Time (Hours)** | **Maximum Inter-Transposition Time (Hours)** | **Minimum Inter-Transposition Time (Hours)** |
|---|---:|---:|---:|---:|---:|
| CNN(1 & 2) | 96 | 7 | 59.12 | 659 | <1 |
| NBC | 16 | 2 | 36.95 | 144 | <1 |
| ABC | 4 | 4 | 18 | 48 | <1 |
| Mixed | 68 | 3 | 46.4 | 321.5 | 1.5 |

Fig. 3.7 illustrates the types (described in Chapter 4) and frequencies of repeated segments. 81% of the repeats are Wild Scenes with no audio; 3% of the repeats are Wild scenes with both audio and video; 14% of the repeats are Comments with both audio and video; and less than 1% of the repeats are Comments with video only and Interviews with both audio and video.

Most of the repeated segments contain only the visual data (i.e., segments shown as a backdrop to a reporter's or an anchor's commentary). Examples include shots of a plane taking off or a missile being fired. Some of the original segments contain comments or a speech in which the source of the audio is

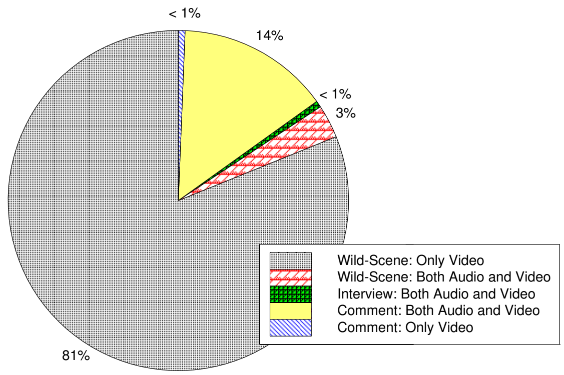


Figure 3.7: Segment Types Repeated by CNN

a subject; however, when the same segment is repeated, the original audio is sometimes suppressed and replaced by a voice-over. For example, initially, a segment of Ms. Albright commenting on Iraq is shown with both video and audio. Later, only the visual is shown with a reporter establishing a context (e.g., "today Albright commented that the situation in Iraq is critical"). Or the visual can be shown as part of field footage (Wild Scene); therefore, no introduction is required.

When a change of context is required, a human editor tries to maintain continuity with an appropriate introduction. However, the temporal continuity is evaluated by assuming that the context is not established before segments with both audio and video are repeated in a composition. In Table 3.10 the repetition of a segment from an earlier time to the future is called flashback and the presentation of a segment from the future without presentation of intermediate information is called a flash-forward. Each time there is a flashback only one segment from the past is repeated; therefore, the segments preceding the flashback segment and the successive segment are in the correct creation time series.

Table 3.10: Temporal Continuity Measurements

| **Parameter** | **Value** |
|---|---|
| Presentation Duration in Hours | 912 |
| Total No. Segs | 387 |
| No. of Segs Repeated (Table 3.9) | 17 |
| Average Inter-Transposition Time in Hours (Table 3.9) | 59 |
| $e_{tc}$ for Flashback | 0.93 |
| Tolerated Forward Jump Value $\delta$ in Hours | 24 |
| $e_{tc}$ for Flash-Forwards | 0.98 |

Temporal continuity between the remaining consecutive pairs is 1 and the mean temporal continuity for the presentation is $1/386(15.8 + 16.6 + 352) = 0.99$.

### Information and Period-Span Coverage

Evaluation of information coverage ($In$) is achieved by an analysis of the information content in the composition set relative to the information contained in the candidate set. Because the contents of the candidate set and the composition set are identical in this case, we do not yield a useful reference for this metric. We also have difficulty with period-span coverage because of the absence of information about the creation time span covered by the original data. Finally, structural continuity is assumed to be inherent in the manually-edited data set and this is consistent with our observations (e.g., CNN rarely makes naive mistakes in assembling video by segment type).

We further observed from the data set that the presentation duration is a varying parameter and its value is highly dependent on the content being presented. When the current focus of the content exhibits changes (i.e., developments and progressions of the event), we observed that the duration of the presentation is longer to support the impact of the content. We also observed that lifespan of news items can vary from a days to years.

Therefore, a candidate set will consist of segments of varying playout duration, period span coverage, and information. These segments need to be selected to form a composition with correct structure, and satisfactory temporal and thematic continuity.

## 3.4 Summary

In this chapter we have formulated a set of metrics to evaluate the quality of a video composition. We consider many essential features of a composition. In particular, we consider information content, information flow, temporal ordering of content, creation time period in a composition, content progression, and structural ordering of information. These features form the essential requirements for formulation of our metrics.

To measure information content in a composition, we compare concepts contained within all the segments in the composition with the concepts in all the segments returned as a result of user-selection criteria (candidate sets). This is achieved by creating centroid vectors of the segments in candidate and composition sets and using cosine metrics to measure the similarity. Thematic continuity, or information flow, is evaluated by measuring similarity in the concepts of the two consecutive segments in a composition. We use cosine metrics for this measurement also. We establish the dissimilarity ($\lambda$) and similarity ($\tau$) thresholds, and consider compositions for which the value of the cosine measurement falls within these thresholds as possessing good thematic continuity.

Temporal continuity is evaluated by using the creation time and date associated with segments in a composition. The quality of a composition with large forward jumps in time or backward jumps in time between consecutive segments is considered poor. Temporal continuity is measured by the difference in creation times of adjacent segments based on a threshold for forward jump $\delta$ in time that can be tolerated in a composition, and a weight $\beta$, for forward jumps. We measure period span coverage by comparing the span covered by segments in the candidate set to the span covered by segments in the composition set.

Content progression in composition is evaluated by measuring the playout durations of constituent video segments. The content progression is considered good if the playout durations fall within the fast-change threshold ($\rho$) and the slow-change threshold, ($\varrho$). Finally, we measure structural continuity using Boolean evaluation.

Utilizing the proposed metrics, we acquired reference values for quality of broadcast news video from CNN, ABC and NBC. Our results show that the thematic continuity varies between 0.50 and 1.0, the low value indicating some rough transitions between consecutive video segments. The content progression varies between 0.81 and 1.0, indicating gradual change in content. The temporal continuity is evaluated to be 0.99. Evaluation of information coverage and period span coverage could not be conducted due to insufficient data. Since the composition is manually composed, the structural continuity is always maintained.

Later in this dissertation (Chapter 5), we will use the above reference values to evaluate composition of a newscast resulting from application of automatic composition and customization techniques proposed in the next chapter (Chapter 4).

# Chapter 4

# Techniques for Composition and Customization of Digital Video

## Synopsis

In this chapter, we present the proposed composition techniques for digital video data. The proposed composition techniques are based on the content within video data, creation time of data, and structure of the video domain. These techniques are applied to news video data. The structure of the resulting composition is based on existing forms or structures for news video composition. The proposed techniques are divided into instance-based and period-based compositions, and include temporal composition, thematic composition, and thematic nearness composition. In addition, we also discuss techniques that make composition under playout time constraints possible.

## 4.1 Introduction

In addition to concepts contained within segments [31, 64, 92], information about creation time associated with the segments and domain-specific structure are also required to produce an automatic video composition. Therefore, we require techniques that consider content, creation time, and structure during a composition. A composition can also be customized, as shown in Fig. 4.1, which depicts composition under playout time constraints (i.e., limited playout duration). Most of the existing customization techniques are based on content customization [47]. However, customization under playout time constraints has not been explored for video data, though trivial playout time constraint techniques (in which a pre-composed presentation is played until the specified duration) are being used to limit the playout time [45]. Thus the challenge is to create a time constraint composition that is cohesive, covers maximum time span of the available information, and presents different aspects of a story.

To demonstrate the composition and customization techniques we use newscast as an example domain; however, the techniques are generic and can be applied to any other domain. We adopt the work of Musburger [63] for correct structural composition of a news item.

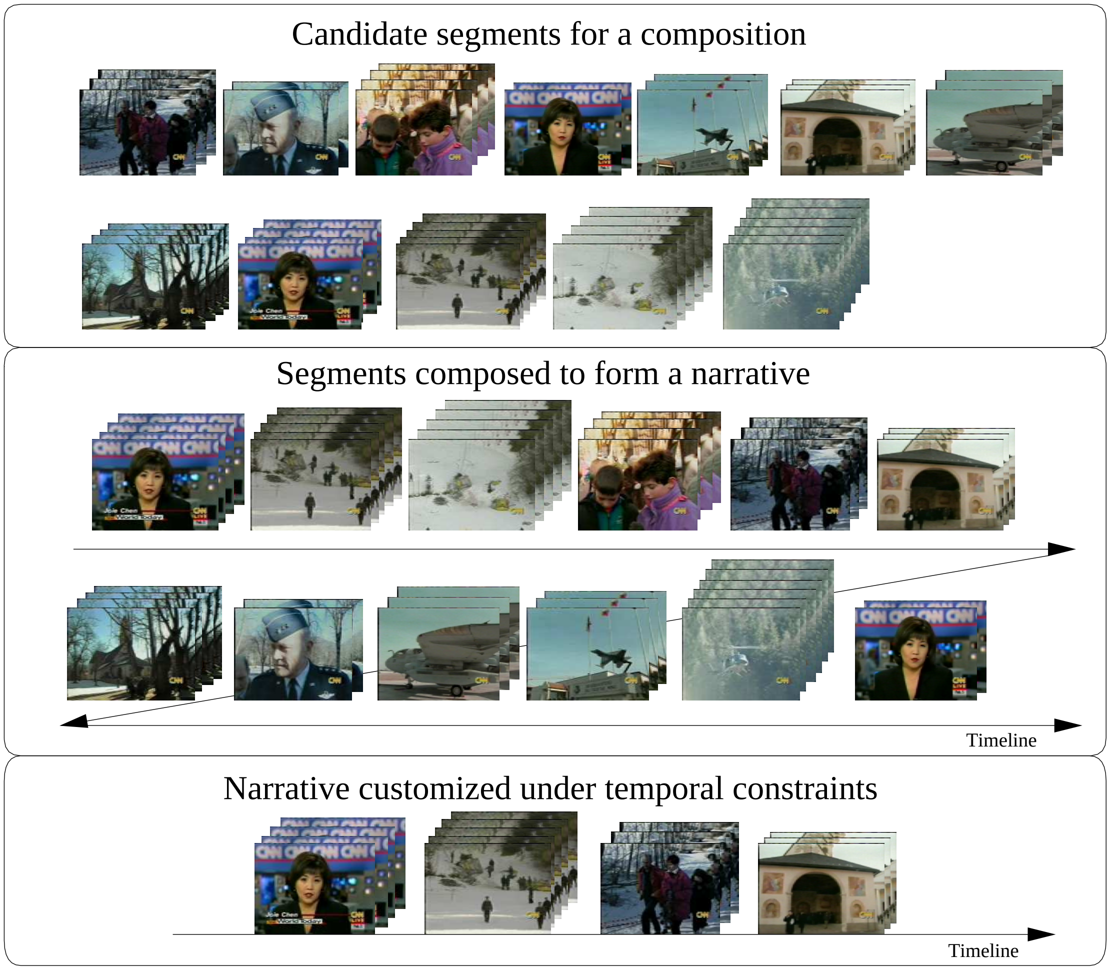

Candidate segments for a composition
Segments composed to form a narrative
Timeline
Narrative customized under temporal constraints
Timeline

Figure 4.1: Schematic of Composition and Customization of News Items

### 4.1.1 Forms of News Items

In a guide to an electronic news gathering, Musburger discusses the various structures or forms associated with a news story. The main forms are as follows:

**Spot News:** Presentation of actuality or scenes of a story that is taking place is called spot news. Usually the story is briefly introduced and the scenes are presented in a linear fashion.

**Stand-Upper:** A reporter gathers information and a videographer shoots as much cover (different threads or perspectives) footage as possible. Finally the story is recorded by the reporter introducing the story followed by the cover footage, and in the end the reporter presents the tag line (ending segment of the story). This form is used in a variety of settings, from breaking news to public relations "puff" pieces.

**Wraparound:** Also called a donut, there are two types of wraparounds. One is the same as a stand-upper, but the anchor delivers the start and the end lines while a reporter delivers the center of the donut. In the second method the center of the donut is an actuality.

As observed from broadcast news and according to Rabiger [75], the ingredients of footage or scenes can be summarized as follows:

**Action Footage or Wild Scenes:** Footage from the actual location of the event is called action footage. Among other things, scenes in an action footage can belong to landscapes, inanimate things, people, or creatures engrossed in everyday activity.

**Interviews:** Interviews refer to one or more people answering formal and structured questions. Interviewers can be off camera and questions can be edited-out.

**Comments:** Informal and on-location interviews with a reporter grabbing someone to interview at the site of the story are referred to as comments.

**Speech:** Speech refers to formal communication or expression of thoughts.

**Re-enactment:** This refers to situations that are already past or cannot be filmed are acted out or animated.

Based on the above forms and footage components we define the structure of a news item. To create a cohesive composition we identify role of segments in a composition and a set of structure-based constraints. In Table 4.1, the structure of a news item is defined as having a beginning, a body and an end. If the segments belonging to a beginning, a body, and an end are transposed, then the news item does not possess structural continuity. The news item should be introduced only once, hence, a news item should start with a single segment of type Introduction, and also end with a single segment of type Enclose. However, the body can have multiple segments depending on the views being presented. If there is no body, then a segment of type Enclose is not included in a composition.

Table 4.1: Structure of a News Item

| | | |
|---|---|---|
| Headline | | |
| Introduction | | |
| Current (body) | Comment | |
| | Wild Scene | |
| | Interview | Question&Answer (QA) |
| | Speech | |
| | Enactment | |
| Enclose | | |

The proposed composition techniques are based on a set of assumptions. We describe them next.

### 4.1.2 Assumptions

To accommodate different news items in a presentation, each news item is allotted a limited duration for presentation. Therefore, the objective of news item composition is to maximize the presentation of the information related to each event in spite of the time constraints. The tactic adopted by newscasters is to provide multiple views of an event rather than a single detailed view. For example, multiple views can include field shots, comments, interviews, and re-enactments of an event after its occurrence. A news item is a collage of views and it is possible to rearrange or drop some of the views to convey the same story.

In the proposed composition techniques we take advantage of the characteristic of news items that are typically short segments that convey a great deal of information (i.e., sound bites). The following assumptions are made during composition:

1. A complete news item is considered an event (e.g., Clinton's visit to South America).

2. An event can be composed of sub-events (e.g., interviews, comments from by-standers, and field shots).

3. Thematic continuity is maintained in a news item when sub-events are presented in an arbitrary order provided:

   (a) all sub-events belong to the same event, and

   (b) each sub-event is completely played-out.

4. Within a news item, all types of segments carry the same theme; however, different types of segments depict the theme from different views and are not redundant. Moreover, these segments are not dependent on one another for presentation.

As defined previously, a presentation possessing thematic continuity is one that comprises segments with related information that are ordered to maintain temporal continuity. Assumption (1) ensures that each news item contains information related to a single event and no irrelevant information is presented. Assumption (2) is required to identify types of segments in a body of an event. Assumption (3) is required to maintain a storyline, or theme, and to avoid abrupt discontinuities in a news item by presenting complete information about each segment. In assumption (4) we treat segments as having content independence so that we can include, exclude or arrange them in any order in the body. A segment is a complete information unit and all dependent content is encompassed in a single segment. For example, if an anchor person introduces a scene, then the introduction is included as part of the scene. Rearrangement of clips is valid only if the segments are from the same instance in chronological time. If available segments are from a period, the flexibility to rearrange them is limited.

Additional symbols used to define news video segment types and the composition techniques are summarized in Table 4.2.

Table 4.2: Additional Symbols Used to Define News Video Segment Types

| **Symbol** | **Description** |
|---|---|
| $S_{sp}$ | Set of single-presentation-type segments |
| $S_{mp}$ | Set of multiple-presentation-type segments |
| $S_{bw}$ | Set of Wild Scene-type segments |
| $S_{bs}$ | Set of Speech-type segments |
| $S_{bi}$ | Set of Interview-type segments |
| $S_{bc}$ | Set of Comment-type segments |
| $S_{be}$ | Set of Enactment-type segments |

## 4.2 Composition Techniques

In this section, we begin with a discussion on the types of segments present in a structure of video-based news. After establishing the basic segment types for this domain, we describe the composition techniques. The taxonomy of Fig. 4.2 illustrates the relationships among the proposed techniques described in this section.

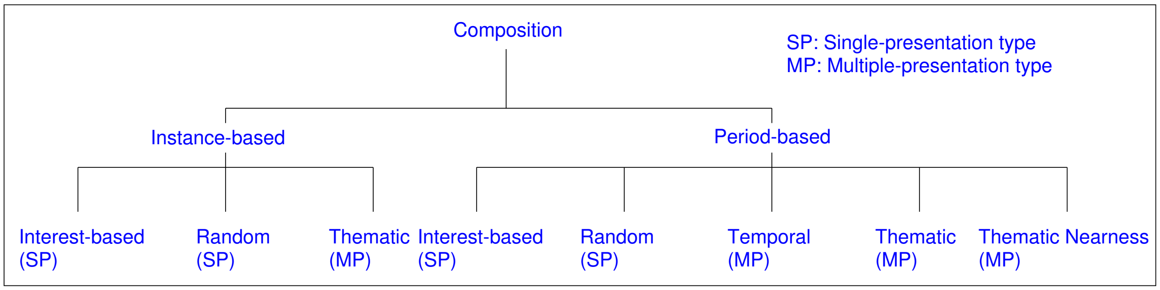


Figure 4.2: Taxonomy of Proposed Composition Techniques

The composition techniques can be divided into two main categories: instance-based and period-based. These are presented in detail in the next section.

### 4.2.1 Segment Types and Structure for the News Domain

We adopt the work of Musburger [63] as a reference structure for composition of a news item. Under this model (Table 4.1), a news item is comprised of an introduction, a body, and an end. Other orderings are invalid and demonstrate poor structural continuity. Moreover, a news item should have a single introduction (segment type Introduction) and a single end (segment type Enclose). However, the body can have multiple segments depending on the views being presented. If there is no body, then a segment of type Enclose is not included in a composition.

Our basic unit of video data in a news item is the segment. However, a segment can also be comprised of multiple segments that form a coherent grouping. For our work a segment can belong to the Comment, QA, Wild Scene, or Enactment types. To conform to these various structures of a news item, we propose a set of rules for composition based on segment type. The types are divided into two categories:

- **Single-presentation type ($S_{sp}$):** The segment types that allow only a single segment of its kind to be included in a composition. This includes segments of type Headline, Introduction, and Enclose.
- **Multiple-presentation type ($S_{mp}$):** The segment types that allow multiple segments of its kind to be included in a composition. This type includes segments that can belong to a body. For example, we can have multiple segments of type Wild Scene in a single news item.

The functions of these categories of segment types are discussed below.

### Single-Presentation Type

To compose a news item we select a single segment of this type. However, the news can be generated from an *instance of creation time* (e.g., today's 7:00 PM news) or over a *period of creation time* (e.g., news about Albright's visit to the Middle East). We use different rules for selection of single-presentation-type segments for instance-based and period-based compositions.

**Creation Instance-Based:** Two techniques can be used to select a segment of the single-presentation-type for the creation instance case. First, selection can be *interest-based*. This is achieved by selecting the segment with the highest selection interest $I(s)$. The segment is defined for a set $S_{sp}$ of the single-presentation-type by the the following predicate:

$$s_k : \exists m : (\forall s \in S_{sp} : m \geq I(s) \wedge m = I(s_k))$$

Second, if all the segments have the same interest value then a *random* selection can be used. A segment $s$ can be selected with a uniform probability. Fig. 4.3 illustrates this type of composition.

**Creation Period:** If a period is indicated, then the rules specified in Table 4.3 are followed to select a single-presentation-type segment.

### Multiple-Presentation Type

Segments of this type belong to the body of a composition. Like the single-presentation-type, the selection of segments is also dependent on instance-based and period-based rules.

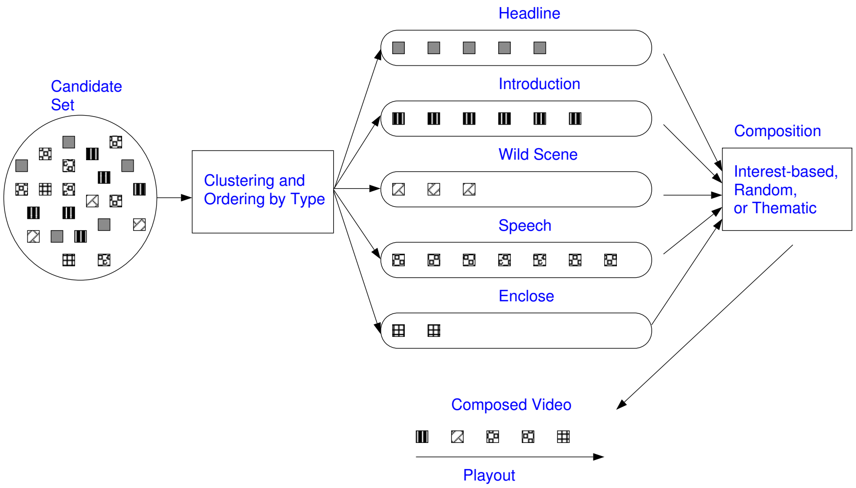


Figure 4.3: An Example of Instance-Based Composition

Table 4.3: Creation Period Composition Rules

| | Rules | Explanation |
|---|---|---|
| 1. | $s_k \| \forall s \in S_h : b_k \leq b \wedge b = b_k$ | To build a news item in chronological order we select a segment belonging to the Headline se the earliest time and date. |
| 2. | $s_k \| \forall s \in S_{in} : b_k \leq b \wedge b = b_k$ | Similarly, we select a segment from the Introduct set with the earliest time and date. |
| 3. | $s_k \| \forall s \in S_e : b_k \geq b \wedge b = b_k$ | We select a segment from the Enclose set that ha the latest time and date. |
| 4. | $s_k \| \exists m : (\forall s \in S_{sp} : m \geq I(s) \wedge m = I(s_k))$ | If more than one segment is available for a parti date then we use the segment $s_k$ with the highes |

**Creation Instance-Based:** Because there can be more than one segment mapping to the same instance on the creation timeline, segments belonging to a body can either be grouped (clustered) depending on their type (i.e., Speech, Interview, Wild Scene, Comment, and Enactment) or not grouped. The reasons for clustering is desire to base a composition on a preference for a particular type or ordering (e.g., Wild Scene before Speech). Another reason is that, for reasons of diversity, the segments are chosen from the different types within the playout time allotment. After forming clusters, the final order of segments in a news item can be determined with the following sequence:

$$[s_h, s_{in}, S_{bs}, S_{bw}, S_{bi}, S_{bc}, S_{be}, s_e],$$

where sets $S_{bs}$, $S_{bw}$, $S_{bi}$, $S_{bc}$, and $S_{be}$ correspond to types Speech, Wild Scene, Interview, Comment, and Enactment, respectively. The order of clusters in a body can be changed based on preference.

**Creation Period:** There are two types of mappings between creation periods and segments contained in a body. Let $s \in S_b$ denote a segment belonging to a body and let an instance of time be represented by $t$. The two types of mappings are then defined as follows:

- One-to-one mapping: The start of a single segment $s \in S_b$ maps to time $t$ (i.e., $s \rightarrow t$) within a period (Fig. 4.4).
- Many-to-one mapping: The start of multiple segments ($\{s_1,\ s_2,\ s_3, ...\} \rightarrow t$) maps to time $t$ (Fig. 4.5).

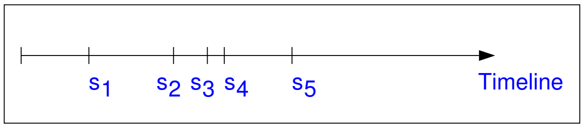


Figure 4.4: An Example of a One-to-One Mapping of Segments to a Timeline

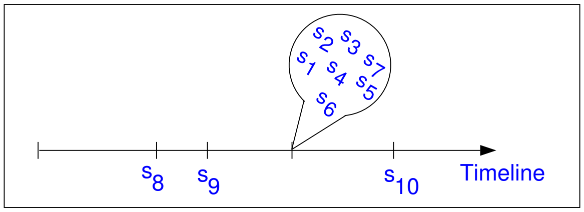


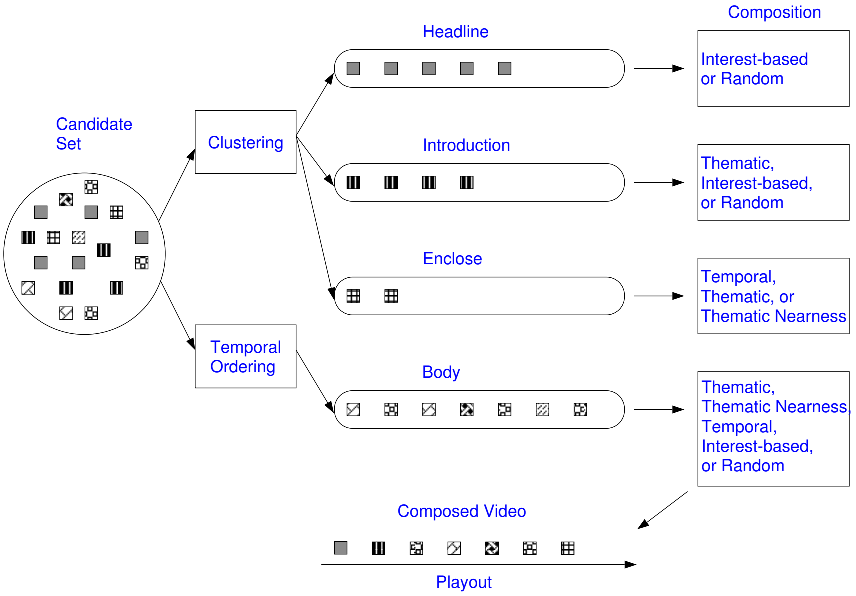
Figure 4.5: An Example of a Many-to-One Mapping of Segments to a Timeline

Segments that map to the same instance are clustered together and are further grouped based on their type – Speech, Interview, Comment, Wild Scene, and Enactment.

After conforming to the structural constraints (Section 3.2), there is still considerable flexibility to select and order segments from the different types. We discuss this next.

### 4.2.2 Techniques for Composition of a News Item

We use interest-based or random selection when there are many single-presentation-type segments that are candidates. However, if the composition is either period-based or requires thematic ordering, then, in addition to the above rules and techniques, we require a strategy to select segments from clusters or from the timeline. Our techniques are based on temporal ordering, temporal continuity, and temporal nearness continuity. This hybrid approach is illustrated in Fig. 4.6, in which clustering and temporal ordering are combined.



Composition
Headline
Interest-based or Random
Candidate Set
Clustering
Introduction
Thematic, Interest-based, or Random
Enclose
Temporal, Thematic, or Thematic Nearness
Temporal Ordering
Body
Thematic, Thematic Nearness, Temporal, Interest-based, or Random
Composed Video
Playout

Figure 4.6: An Example of Hybrid Composition

### Temporal Ordering

This scheme is applicable to period-based compositions. Segments are organized on the timeline as a chronology according to their creation time and date (Fig. 4.7).

In this case the single-presentation-type segments are selected according to the rules of Table 4.3. The resulting composition set consists of a single seg-

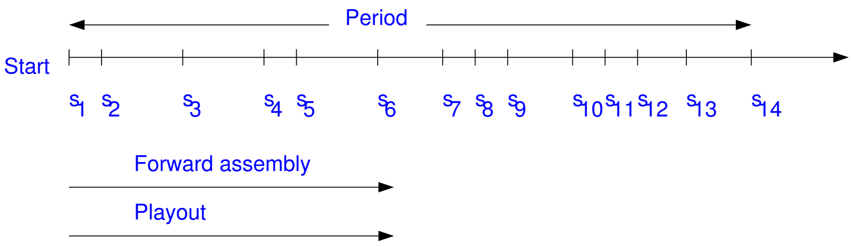


Figure 4.7: Forward Temporal Ordering Scheme

ment of each single-presentation-type in $S_a$ and all multiple-presentation-type segments in $S_a$. To achieve composition, the segments in $S_a$ are sequenced using structural constraints in increasing order of creation time and date. The objective of this technique is to obtimize the information in the presentation (i.e., include all possible segments in the final composition),temporal continuity, and target span covered.

In the above composition we assume that all of the candidate segments belong to the same story center and the segments are created as the event evolves. Continuity is provided by temporal ordering. The segments selected to compose a news item contain information related to a story center. However, since a news item develops over time and there are variations in theme due to multiple threads of the story, lower thematic continuity results. An extension to this technique, thematic composition, aims to ensure that there are no large jumps in themes between consecutive segments due to these threads.

### Thematic Composition

For temporal ordering we depend on the simplicity of the ordering among the segments to provide thematic continuity. However, as evident from the characteristic of conventional news video, a composition can be acceptable with other types of orderings yielding different thematic continuities. Therefore, we try to achieve composition of segments with related information with an ordering that maintains temporal continuity. We use *concept similarity* ($CS$) for the sequencing.

The concept similarity between two segments can be found by using the cosine similarity metric. The composition begins by selecting the first segment of the single-presentation-type (Headline or Introduction) using interest-based or random selection. If a Headline segment $s_h$ is selected then an Introduction segment $s_{in}$ is selected using the concept similarity. Otherwise, if $s_{in}$ is selected first, then interest-based or random selection is used.

The first body-segment is selected by considering its concept similarity with $s_{in}$. Next, after the first body-segment on the timeline has been selected, the proceeding segment is included or dropped depending on the similarity and dissimilarity thresholds, $\tau$ and $\lambda$. If the similarity value of two segments $d(s_i, s_j)$ is more than $\tau$, then the two segments are considered to be the same and only one is used in the composition. If the similarity value of two consecutive segments from $S_a$ is less than $\lambda$, then the two segments are not considered similar. When all pairs are exhausted, Eq. 4.1 is valid for all consecutive segments $s_i$ and $s_j$ in the composition:

$$\lambda \leq \text{cosine}(\vec{W}_i, \vec{W}_j) \leq \tau. \tag{4.1}$$

There are different requirements for selecting segments for an instance-based or period-based scenario. These are discussed below.

**Creation Instance:** When building a composition for a creation instance, we begin by selecting a Headline or an Introduction segment according to the interest-based or random technique. If a segment of type Headline is selected to start the composition, then the next segment of type Introduction is selected based on concept similarity. However, if there are multiple segments with the same concept similarity, then the final segment selection is interest-based or random. If the instance-based segments are not already clustered, then they are sequenced by finding the concept similarity among them. If the segments are grouped according to their type, then segments in the first group are sequenced based on concept similarity and then incorporated in the composition. Likewise, segments from the next group are sequenced and incorporated until all groups are sequenced and incorporated in the composition. Concept similarity is maintained between the groups by selecting the first segment from the proceeding group that is similar to the last segment sequenced from the preceding group. Some multiple-presentation type segments from the groups need not be incorporated in order to maintain concept similarity.

The Enclose segment is again selected based on concept similarity; however, if there are multiple segments with the same concept similarity then the final segment selection is interest-based or random.

**Creation Period:** For a creation period composition, all segments belonging to the body are chronologically ordered initially so that the predicate in Eq. 3.1 holds for all segments. The rules for selection of single-presentation-type segments (Table 4.3) and the selection of segments from clusters of multiple-presentation-types are the same as for compositions of a creation instance; however, the generated composition must be valid for the predicate of Eq. 3.1.

Under these conditions, the objective of the final composition is to optimize the thematic continuity. Although such a composition can possess large forward time discontinuities and loss of information $In$. This is evident in the analysis of Section 5.4.

### Thematic Nearness Composition

We introduce thematic nearness in order to achieve good thematic continuity but without the large temporal discontinuities associated with the thematic composition technique. This technique also reduces the probability of incorporating only a single thread into the composition. To achieve this, we observe that segments along a timeline belonging to the same thread have a high level of similarity even as the thread progresses. Information similarity is a function $IS$ of concept similarity ($CS$) and the difference ($b_i - b_j$) in creation time and date between segments $s_i$ and $s_j$.

$IS$ is directly proportional to $CS$ ($IS \propto CS$) (i.e., similarity between two segments increases with the number of common segments). $IS$ is inversely proportional ($IS \propto \frac{1}{b_i - b_j}$) to distance ($b_i - b_j$) on the timeline (i.e., segments with similar information must be closer in creation time). For maintaining thematic continuity, successor segments are created at the same time or later than their predecessors. Therefore, for any sequential $i$ and $j$ the value of $b_i - b_j$ should be positive. $IS$ between segments is defined as:

$$IS(s_i, s_j) = A \times \frac{\text{cosine}(\vec{W}_i, \vec{W}_j)}{b_i - b_j}, \tag{4.2}$$

where $A$ is a normalization constant used for convenience. We assume uniform distribution of segments along the timeline. If $(b_i - b_j) = 0$, (i.e., more than one segments maps to the same time) and use the cosine metric for measurement of similarity between the two segments. This type of composition will result in lower relative thematic continuity, and will reduce the occurrence of dropped segments or temporal discontinuities. Therefore, the objective of this composition is to simultaneously optimize information, thematic continuity, temporal continuity, and target span covered. This is evident in the evaluation in Section 5.5. Fig. 4.8 illustrates the relationships among temporal, thematic, and thematic nearness compositions.

The above composition approaches allow a selection of segments to achieve target goals such as thematic continuity. However, additional techniques are required to deal with constraints on the duration of the final composition.

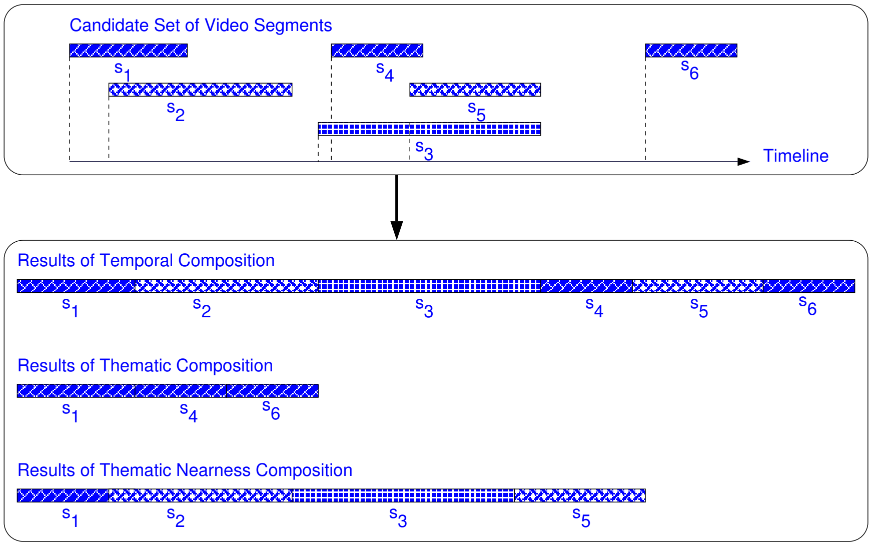

Candidate Set of Video Segments
s1
s2
s3
s4
s5
s6
Timeline
Results of Temporal Composition
s1
s2
s3
s4
s5
s6
Results of Thematic Composition
s1
s4
s6
Results of Thematic Nearness Composition
s1
s2
s3
s5

Figure 4.8: Example Illustrating Relationships Among Composition Techniques

### 4.2.3 Composition Under Time Constraints

We base our approach for composition under time constraints on two assumptions. First, we assume that each segment type in the body presents information about an event from a different aspect. Second, we assume that each segment in the body is independent of the others. The implications of these assumptions are that discarding segments from the body of a news item or including segments in the body from various sources will not substantially degrade thematic continuity of the composition. We consider scenarios for the composition of single and multiple news items under a time constraint.

When there is ample time for the set of composed segments additional content can be selected to augment the composition (single or multiple news items). When there is insufficient time, we must drop or cut some of the segments to fit the constraint. Let $d_u$ specify the target composition duration and $d_c$ represent the time required for the overall composition (single or multiple items). For a composition with single item $d_c$ is reduced to $d_{S_c}$. The two cases are considered below.

### Insufficient Time Case

When there is insufficient time to accommodate the complete composition sets we must drop some segments. If we use the thematic composition technique, dropping can be achieved by decreasing the value of $\tau$ so that additional segments are considered to have the same content and are eliminated from the composition. A similar result can also be achieved by increasing $\lambda$. By using this approach, fewer threads are encompassed and the information level ($In$) of the composition decreases.

Another approach is to distribute the available duration across the composition sets. In this case each item gets an equal opportunity to be part of the complete composition. However, this can result in incomplete composition of individual items.

The structural-based temporal exclusion rules of Table 4.4 are used to form complete and cohesive news items. These rules dictate the time allocated for each item while preserving cohesion.

If the application of these techniques fails to reduce the composition set duration to within the constraint then we seek to drop segments from within the domain-specific components. For news video we look to drop seg-

Table 4.4: Exclusion Rules for Time-Constrained Composition

| | Rules | Explanation |
|---|---|---|
| 1. | $(d_u < d_c) \Rightarrow ((s \in S_h) \not\subset S_c)$ | If the duration of a news item is less than required, then the segment of type Headline is dropped. |
| 2. | $(d_u < d_c) \Rightarrow ((s \in S_e) \not\subset S_c)$ | After dropping the headline, if the duration of a news item is still less than required, then the segment of type Enclose is dropped. |

ments from the body of the news items. We propose heuristic techniques for instance-based and period-based compositions.

**Creation Instance Adjustments:** For the creation-instance case, we attempt to incorporate the greatest diversity of segment types into the composition at the expense of the depth of each segment type. This is a typical knapsack problem [29], the objective of this type adjustment algorithm is to optimize views or information and utilize as much of the available playout duration as possible. For example, if there are multiple speech segments that cannot all be accommodated then initially only one is selected. Similarly, a single question and answer can be selected to comprise an interview segment. This process continues until all of the content is spanned. The number of components in the composition increases with each pass. The associated Creation-Instance Adjustment Algorithm leads to composition under playout time constraints. The algorithm, shown below, takes a composed sequence (e.g., a news item) and an allocated duration $d_{S_c}$ as input and produces a modified set $S_c$. The segments can be re-sequenced for presentation.

**Creation-Instance Adjustment Algorithm:**

```
1 Select the Introduction
2 If the Introduction segment duration is less than or equal to the allocated time $d_{S_c}$ then
  2.1 Decrease the allocated time by the current segment duration ($d_{S_c} \leftarrow d_{S_c} - d_s$)
  2.2 For each unvisited segment in all groups and allocated time remaining
      2.2.1 For each group type in the body and allocated time remaining
      2.2.1.1 For each segment in the group and no segment selected from the group
          2.2.1.1.1 If the duration of the segment is less than or equal to the allocated time then
              2.2.1.1.1.1 Select the segment for the composition
              2.2.1.1.1.2 Decrease the composition duration by the duration of the current segment
  2.3 If an Enclose segment available and its duration is less than or equal to the allocated time then
      2.3.1 Select the segment for the composition
3 Else end (the composition does not fit the time allocation)
```

The example in Table 4.5 illustrates a composition set for one news item. The application of the rules on this composition set with a target duration of 600 seconds yields the result: Introduction, Speech$_1$, Speech$_2$, Wild Scene$_1$, Comment$_1$, Wild Scene$_2$, QA$_{11}$, QA$_{21}$.

The duration $d_{S_c}$ is allocated proportional to the complete playout time of a composition. Hence, if there are $k$ compositions in a collection and there are $n$ segments in each composition, then each composition is allocated a duration:

$$d_{S_{ci}} \leftarrow \frac{\sum_{m=1}^{n} d_m}{\sum_{j=1}^{k} \sum_{l=1}^{n} d_{s_{il}}} \times d_u. \tag{4.3}$$

Table 4.5: Example of Creation Instance Time Adjustment

| Introduction | Body | | | | Enclose |
|---|---|---|---|---|---|
| Introduction(10) | Speech$_1$(60) | Wild Scene$_1$(30) | Interview$_1$<br>QA$_{11}$(60)<br>QA$_{12}$(130)<br>QA$_{13}$(100) | Comment$_1$(20) | Enclose(22) |
| | Speech$_2$(180) | Wild Scene$_2$(40) | Interview$_2$<br>QA$_{21}$(200)<br>QA$_{22}$(50) | Comment$_2$(15) | |
| | | Wild Scene$_3$(14) | | Comment$_3$(9) | |

Some of the allocated duration can still remain for each composition in a collection as playout duration of a segment cannot fit the available allocated duration. At this stage the knapsack problem is simply reduced to the bin packing problem [27], the remaining un-allocated durations from the compositions in a collection are accumulated and we try to fit segments from the compositions into the accumulated remaining time, such that, on average, least amount of time is left un-utilized. At this stage we do not care about the views (information) contained in the segments. Therefore, we optimise the playout duration of a composition.

To best fit a segment into the remaining time, we analyzed the following schemes:

**Best Fit Across all Compositions (BFAC):** Select the segment with the largest playout durations across all compositions and then select the segment with the second smallest and so on. The process continues until all the remaining time is used up or the playout durations of segments that have not been selected are larger than the remaining time.

**Least Fit Across all Compositions (LFAC):** Select the segment with the smallest playout durations across all compositions and then select the segment with the second smallest and so on. The process continues until all the remaining time is used up or the playout durations of segments that have not been selected are larger than the remaining time.

**Best Fit in a Composition (BFIC):** Select the segment with the largest playout duration in a composition that has not been already selected. Iterate through all the compositions selecting the segment with largest playout duration that has not been already selected until all the remaining time is used up or the playout durations of segments that have not been selected are larger than the remaining time.

**Least Fit in a Composition (LFIC):** Select the segment with the smallest playout duration in a composition that has not been already selected. Iterate through all the compositions selecting the segment with smallest playout duration that has not been already selected until all the remaining time is used up or the playout durations of segments that have not been selected are larger than the remaining time.

**First Come First Select in a Composition (FCFS):** Select the first segment in a composition that has not been already selected. Iterate through all the compositions selecting the first segment that has not been already selected in a composition until all the remaining time is

Table 4.6: Performance of Bin Packing Schemes

| Average $d_u$ | Time Leftover (Seconds) | | | | |
|---|---|---|---|---|---|
| (Seconds) | BFAC | LFAC | BFIC | LFIC | FCFS |
| 327 | 4.22 | 11.09 | 6.06 | 7.45 | 6.82 |

Table 4.7: Effect of Bin Packing Schemes on $e_{thc}$ & $e_{cp}$

| $e_{thc}$ | | | $e_{cp}$ | | |
|---|---|---|---|---|---|
| BFAC | BFIC | FCFS | BFAC | BFIC | FCFS |
| 0.96 | 0.94 | 0.98 | 0.92 | 0.93 | 0.91 |

used up or the playout duration of segments that have not been selected is larger than the remaining time.

Table 4.6 illustrates the performance of the about five schemes with respect to the time leftover in a collection of composition after bin packing. The observations are based on 100 queries with varying $d_u$.

Clearly the LFAC and LFIC schemes did not perform well. Next, we evaluated the effect on thematic continuity and content progression of 375 compositions using BFAC, BFIC, and FCFS schemes as shown in Table 4.7.

As expected the thematic continuity of the BFAC scheme is not as good as the thematic continuity of the FCFS scheme. However, the performance of the BFAC scheme is better then the BFIC scheme as the number of segments selected by the BFAC scheme are less and hence, spoils the thematic continuity of fewer compositions. The content progression of the FCFS scheme is more depictive of the content progression of the original compositions. However, the content progression of the BFAC is worse than BFIC as the content progression becomes more static by selecting the largest segment.

Comparing the overall features (i.e., $e_{thc}, e_{cp}$, and leftover time) the BFAC scheme performs best and we implement this scheme in our algorithms. The steps for accommodating a collection of creation-instance-based compositions under the time-limited constraint are as follows:

```
1.  Allocate time $d_{S_{ci}}$ proportionately to all compositions
2.  Use creation-instance adjustment algorithm for each composition
3.  Accumulate remaining times from all compositions
4.  If accumulated time is greater than zero then
    4.1  Try to accommodate all the compositions (by selecting introduction) that could not be selected
         in the step 2
5.  Use BFAC scheme to bin pack segments if any time remains
```

**Creation Period Adjustments:** For the creation-period case, we attempt to incorporate segments from most of the creation period. This is also a knapsack problem [29], the objective of this adjustment algorithm is to optimize the target span covered and utilize the available playout duration as much as possible. We divide a creation period into sub-periods $TP_i$ to differentiate segments on the timeline. Fig. 4.9 shows a creation timeline divided into periods of 24 hours (e.g., 24, 48, 72, 96). All segments are chronologically ordered on the creation timeline. Segments comprising a composition resulting from any period-based composition techniques are used for playout time adjustment.

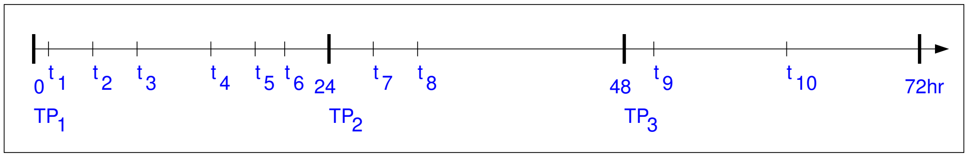


Figure 4.9: Dividing Periods for Temporal Constraint Composition

If the constraint duration is less than the total time of the composition set then segments from some periods must be dropped. This can be achieved by forward or reverse assembly. Forward assembly selects items from the start of each period. Once the available time is consumed then subsequent sub-periods cannot be assembled. Reverse assembly selects items from the end of the sub-period, working backwards in time. When time runs out then the earlier sub-periods cannot be adjusted.

For forward assembly, we use a forward breadth-first and depth-second approach. Assume that a playout period for a news item $TP$ consists of $\{TP_1, TP_2, ...,$

$TP_n\}$ sub-periods as shown in Fig. 4.9. Staring with the first sub-period $TP_1$, we compose a body by selecting one segment from each sub-period per iteration. After each iteration, if time is left then we select additional segments by visiting the sub-periods again until all of the time has been adjusted. Selection from each sub-period is performed in chronological order. If a cluster of segments (belonging to the body) mapped to an instance is encountered, then only a single segment is selected from the cluster per iteration. After all possible one-to-one segments in the sub-periods are accommodated and there is still time left, we then revisit the clusters (many-to-one mappings) and try to adjust the content from them. Each cluster from each sub-period $TP_i$ is visited in chronological order.

The rules of Table 4.4 and the Creation-Period Adjustment Algorithm, shown below, are applied to achieve these results. Segments mapping to an instance in period-based customization can also be incorporated using an instance-based breadth-first and depth-second approach. Similarly, we can use a reverse breadth-first and depth-second approach. In this case we begin composition from the last sub-period. However, the segments are composed to appear in chronologically-ascending order. The algorithm also takes a composed sequence (e.g., a news item) and an allocated duration $d_{S_c}$ as inputs and produces a modified set $S_c$. The set is re-sequenced as a chronology for presentation.

**Creation-Period Adjustment Algorithm:**

```
1   Select the Introduction
2   If the duration of the Introduction segment is less than or equal to the allocated time then
    2.1   Decrease the allocated time by the current segment duration
    2.2   For each unvisited segment in all sub-periods and allocated time remaining
          2.2.1   For each sub-period in the body and allocated time remaining
                  2.2.1.1   For each segment in the sub-period and no segment selected
                            2.2.1.1.1   If a single segment is mapped to time $t$ in a sub-period then
                                        2.2.1.1.1.1   If the duration of the segment is less than or equal to the allocated time then
                                                      2.2.1.1.1.1.1   Select the segment
                                                      2.2.1.1.1.1.2   Decrease the allocated time by the current segment duration
                            2.2.1.1.2   If multiple segments are mapped to time $t$ in a sub-period then
                                        2.2.1.1.2.1   For each segment in the group and no segment selected
                                        2.2.1.1.2.1.1   If the duration of the segment is less than or equal to the allocated time then
                                        2.2.1.1.2.1.1.1   Select the segment
```

```
            2.2.1.1.2.1.1.2  Decrease the allocated time by the current segment duration
    2.3  If an Enclose segment is available and its duration is less than or equal to the allocated time then
        2.3.1  Select the segment for the composition
3  Else end (no composition fits)
```

Similar to the creation-instance adjustment algorithm, in creation-period adjustment algorithm the user specified duration $d_u$ is apportioned among the compositions using Eq. 4.3. In a creation-period-based composition there should not be large forward jumps or else the temporal continuity of a presentation can be compromised. Therefore, for bin packing, the BFAC scheme cannot be used across all sub-periods of a composition but is used across all sub-periods $n$ (that have been already spanned by the creation-period adjustment algorithm) in all compositions that have already been spanned by the creation-period adjustment algorithm. If no segment is selected, we try to accommodate segments from the $n+1th$ sub-period (the next sub-period that has not been spanned) across the composition, if time still remains we give up adjusting time as not to reduce the temporal continuity of a composition.

The steps for accommodating a collection of creation-period-based compositions under the time-limited constraint are as follows:

```
1  Allocate time $d_{S_{ci}}$ proportionately to all compositions
2  Use creation-period adjustment algorithm for each composition
3  Accumulate remaining times from all compositions
4  If accumulated time is greater than zero then
    4.1  Try to accommodate all the compositions (by selecting introduction) that could not be selected
         in the step 2
5  Use BFAC scheme to bin pack segments
```

**Window-based composition:** We can also specify composition to be based on a fractional use of the available composition set. For example, one might specify the selection of 20% of the available content (50 minutes), yet require this to be rendered in a constrained duration of 10 minutes. Three types of window mappings for this selection are proposed:

1. **Start-map window:** The start of the window coincides with the start of the period for which we have data available. In a start-map window the stop point is defined beforehand. This yields a composition based on the earliest available content.

2. **End-map window:** The end of the window coincides with the end of the period for which we have data available. In a end-map window the stop point is defined beforehand. This yields a composition based on the most recent available content.

3. **Middle-map window:** The start and end of the window coincides with a portion of the period for which we have data available.

In each case, creation-period adjustment algorithm can be used for composition.

### Ample Time Case

If $d_u > d_c$, then the complete set of segments can be accommodated. However, there is unused time available for the final composition. To consume this leftover, we can select related unused content from the associated candidate sets. In the news domain, Wild Scenes, when available, and if already selected in a composition, can be repeated as filler. The leftover time $(d_u - d_c)$ is divided proportionally among the compositions in a collection. The playout duration apportioned to a composition is based on the total playout duration of segments that can be used as fillers in a composition. If there are no segments in a composition that can be used as fillers, the composition is not considered for filling. The Filler Algorithm shown below leads to the accommodation of the leftover time $(d_u - d_c)$. The algorithm takes a composed sequences (e.g., news item) and the leftover time and produces composed sequences augmented by additional segments in the body of the composition.

**Filler Algorithm:**

```
1   For each candidate segment in the composition and a nonzero leftover time
    1.1   If the segment duration is less than or equal to the leftover time then
          1.1.1   Include the segment in the composition
          1.1.2   Decrease the leftover time by the included segment duration
2   If the leftover time is greater than zero then
    2.1   If Wild Scene segment not already selected as filler exists then
          2.1.1   Select the partial segment with playout duration equal to the leftover time
3   End
```

After using the filler algorithm once, it is a possibility that there is some time leftover in a composition; for example, all wild scenes are used as fillers and time is still left. The time left from each composition in a collection can be accumulated and used to fully accommodate the partial segments. The steps for filling a collection of compositions are as follows:

```
1   Allocate time $d_{S_{c\,i}}$ proportionately to all compositions
2   Use filler algorithm on each composition
3   Accumulate remaining time from all compositions
    3.1   If accumulated remaining time is greater than zero then
          3.1.1   Try to fully accommodate all partial segments in the compositions
```

A recognized problem with this approach is the fragmentation due to the introduction of incomplete segments. An alternative approach is to introduce a completely different type of filler such as advertisements.

## 4.3 Summary

In this chapter, we have presented the proposed composition and customization techniques for producing a video piece from related segments. Our proposed techniques are based on the existing structures of news composition, and produce video that possesses correct time series of concepts or threads, and smooth flow of theme.

The segments used in a composition are divided into two types: single-presentation type and multiple-presentation type. The single-presentation type include segments belonging to type Headline, Introduction, and Enclose. The multiple-presentation type include segments belonging to type Wild Scene, Comment, Interview, Speech, and Enactment. Only a single segment from any type belonging to the single-presentation can be present in a composition. However, multiple segments from any type belonging to the multiple-presentation can be present in a composition.

Our composition techniques are divided into two types: instance-based and period-based. In instance-based composition, the creation time of all the segments map to a single instance on chronological timeline, and hence there is a many-to-one mapping between the segments and the timeline. In a period-based composition the creation time of the segments map to a different instance on chronological timeline, and hence there is a one-to-one mapping between the segments and the timeline (in some cases more than one segment can map to the timeline).

In instance-based composition, we assume that all the data share the same concepts. Hence, segments in the body of a composition can be randomly ordered with little or no loss in thematic continuity. Our thematic composition technique can be used to obtain better information flow in the composition.

In period-based composition we include temporal, thematic, and thematic nearness ordering techniques. In temporal composition, we assume that a simple ordering of the segments according to their creation time and structural specification will lead to a cohesive composition. However, temporal ordering will not result in a composition with good thematic continuity because a storyline possesses a number of threads that offer different views and there can be a variation in information among threads. We use thematic composition to address this weakness.

In thematic composition, we use structural and temporal ordering; in addition, we select segments based on the concept similarity between two segments. We use cosine metrics to measure concept similarity between the two segments. However, since the resulting composition has a tendency to stick to a small number of threads, and consequently result in large temporal jumps, we use a thematic nearness technique to reduce temporal jumps. In this technique, similarity between segments is evaluated based on the concepts within the segments and normalized with the difference of creation time between the two segments. Segments with a smaller number of common concepts but with relatively closer creation time result in higher similarity value than segments with larger number of common concepts but relatively far apart in creation time. The resulting composition consists of a larger number of threads and smaller temporal jumps, and possesses better thematic continuity than the temporal composition technique but lower thematic continuity than the thematic composition technique.

Our proposed techniques for composition under playout time constraints are based on the assumptions that each kind of segment (e.g., Wild Scene, Interview, and Comment) presents information from different aspects and each segment in the body is independent of the other for purposes of presentation. We present information from many different aspects and cover as much of the creation time period as possible. For the creation-instance time-limited composition technique we incorporate the diversity in the types of segments as opposed to the number of segments of each type. In the creation-period time limited composition technique we divide the complete period into sub-periods and incorporate segments from as many sub-periods as possible as opposed to the number of segments from each sub-period.

In the above techniques, if a presentation consists of more than one composition, then the specified playout time is distributed among all the compositions. Each composition is assigned a time proportional to its total playout duration. Any left over time from the compositions is accumulated and bin packing technique (BFAC) is used to fill up the time. In this technique, the segment with the largest playout time (not already included) is first selected, followed by the second largest segment, and so on, until no more segments can be accommodated.

# Chapter 5

# Evaluation of the Proposed Composition and Customization Techniques

## Synopsis

In this chapter, we use the metrics proposed in Chapter 3 to quantify the quality of a composed news video piece resulting from the proposed composition techniques (Chapter 4). We evaluate the quality of video pieces composed by using temporal, thematic, and thematic-nearness techniques. We also evaluate the effect on quality of playout-time-constrained video compositions using the time limited techniques.

## 5.1 Introduction

News items/events in a broadcast news session are sequenced according to their importance. We observed that the greater the importance of a news item the earlier it is presented in a session. Further, the duration of the presentation of a news event depends on the importance of the content or sub-event being presented. In Fig. 5.1 broadcast durations of a single instance/day of a single news item from the three sources are shown. The relative (within a source) extreme variations in presentation duration are a function of the content importance (e.g., in a murder story if the culprit is caught then more time is given to the news item).

In a period-based composition, if the playout duration of a news item is not constrained, then all related data are composed. Therefore, in the analysis presented in this chapter, the quality of automatically composed news items that are much longer than conventional broadcast news items is also evaluated.

We present an analysis of news items composed using instance-based composition in which we analyze the quality of a news item when segments are shuffled. We present the analysis of a period-based temporal, thematic, and thematic nearness compositions. For thematic and thematic nearness composition we evaluate a number of compositions of the same single news item. The different compositions are achieved by varying the value of dissimilarity threshold $\lambda$. By changing the value of $\lambda$ we demonstrate the concept of thread inclusion in a composition. As the value of $\lambda$ increases, the thematic

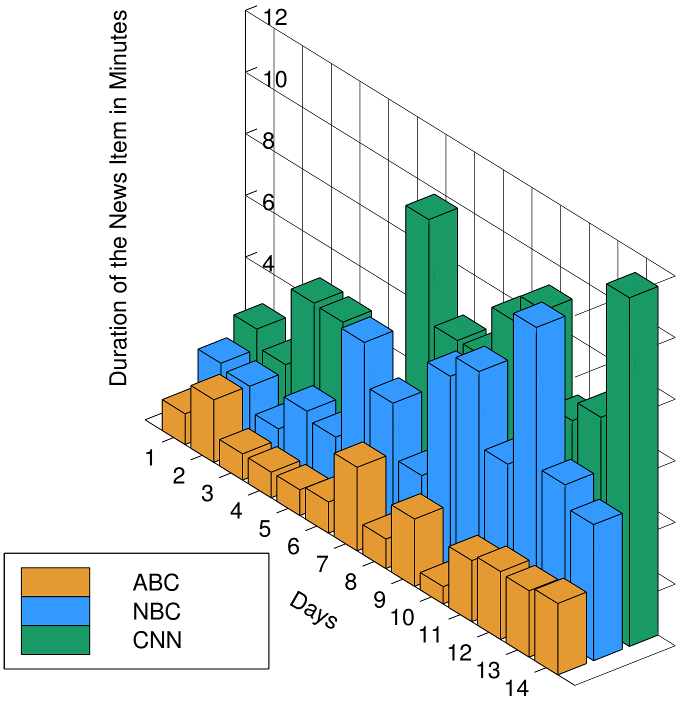


Figure 5.1: Delineation of Broadcast Durations of a Single News Item from Different Sources Over a Period of 14 Days

jump (Fig. 3.1) decreases, hence, including a smaller number of threads in a composition.

Note that we do not evaluate structural continuity here because we expect structural constraints already to be enforced, resulting in a structural continuity equal to one.

We have 10 hours of digitized news video data and their corresponding closed-caption data acquired from the network sources. The data set contains 335 distinct news items obtained from CNN, CBS, and NBC. The news items comprise a universe of 1,731 segments. The playout duration $d$ or the segments varies between 2 seconds and 140 seconds.

To evaluate the composition techniques we use data from four news topics: "United Nations and Iraq Standoff," "Clinton and Intern Controversy," "The Pope's Visit to Cuba," and "Alabama's Bombing Incident." Data for these news topics cover a period of two to fifteen days. The performance of content progression in the results represent the playout duration of segments in the original broadcast composition.

## 5.2 Instance-Based Composition

In this section we analyze the effect of shuffling segments that are mapped to an instance on a historical timeline. In Table 5.1 we show the thematic continuity of the segments sequenced in actual broadcast news. We then shuffle these segments around. The segments in a broadcast news item are assumed to be from the same instance on the creation timeline and hence contain information from the same sub-event. However, news items that contain transposed segments from previous instances have lower thematic continuity. The segments are clustered according to their type and ordered as follows:

$$\text{Wild Scene} \rightarrow \text{Comment} \rightarrow \text{Interview} \rightarrow \text{Speech} \rightarrow \text{Enactment}$$

As seen in Table 5.3, the thematic continuity before and after the segment shuffling does not deteriorate considerably and remains within the range of reference thematic continuity values. In some cases the thematic continuity improves.

## 5.3 Temporal Ordering

Table 5.4 shows the behavior of the period-based temporal ordering technique applied to the segments in the body of a composition. In this technique we simply order the segments along a timeline. As a result, temporal continuity is highly dependent on the default continuity among the segments. During measurement of the temporal continuity the tolerated value of a forward jump, $\delta$, is assumed to be 24 hours. Because all available data are composed in these compositions, the value of the information and period span metrics are equal to one. We also ensure that segments are not repeated or transpositioned here and any degradation in temporal continuity is then due to large forward temporal spans between consecutive segments if the segments only exist for those temporal spans.

Table 5.1: Evaluation of Instance-Based clustered Composition

| **Comp. #** | **No. of** | $e_{tc}$ | $e_{tc}$ | **Comp. #** | **No. of** | $e_{tc}$ | $e_{tc}$ |
|---:|---:|---:|---:|---:|---:|---:|---:|
| | **Segs** | **Broadcast** | **Clustered** | | **Segs** | **Broadcast** | **Clustered** |
| 1 | 9 | 0.97 | 0.93 | 39 | 4 | 1.00 | 1.00 |
| 2 | 9 | 0.91 | 0.46 | 40 | 6 | 1.00 | 1.00 |
| 3 | 8 | 0.97 | 0.94 | 41 | 10 | 1.00 | 0.98 |
| 4 | 7 | 1.00 | 0.94 | 42 | 7 | 1.00 | 1.00 |
| 5 | 8 | 0.89 | 0.90 | 43 | 5 | 1.00 | 1.00 |
| 6 | 7 | 0.89 | 0.75 | 44 | 5 | 0.98 | 0.98 |
| 7 | 10 | 0.93 | 0.70 | 45 | 7 | 1.00 | 1.00 |
| 8 | 6 | 0.94 | 0.98 | 46 | 4 | 1.00 | 1.00 |
| 9 | 10 | 0.98 | 0.92 | 47 | 5 | 0.97 | 0.97 |
| 10 | 6 | 0.99 | 0.97 | 48 | 5 | 1.00 | 1.00 |
| 11 | 7 | 1.00 | 1.00 | 49 | 10 | 0.99 | 0.98 |
| 12 | 5 | 1.00 | 0.98 | 50 | 8 | 0.98 | 0.98 |
| 13 | 6 | 0.95 | 0.89 | 51 | 5 | 0.94 | 0.90 |
| 14 | 8 | 0.97 | 1.00 | 52 | 5 | 0.89 | 0.87 |
| 15 | 9 | 0.99 | 0.99 | 53 | 4 | 1.00 | 1.00 |
| 16 | 7 | 0.97 | 1.00 | 54 | 6 | 1.00 | 1.00 |
| 17 | 4 | 1.00 | 1.00 | 55 | 7 | 1.00 | 0.99 |
| 18 | 7 | 0.98 | 0.99 | 56 | 5 | 0.98 | 0.98 |
| 19 | 14 | 1.00 | 1.00 | 57 | 11 | 1.00 | 1.00 |
| 20 | 12 | 0.99 | 1.00 | 58 | 6 | 0.98 | 0.96 |
| 21 | 4 | 0.94 | 1.00 | 59 | 4 | 1.00 | 0.99 |
| 22 | 12 | 1.00 | 1.00 | 60 | 4 | 1.00 | 0.99 |
| 23 | 11 | 0.98 | 0.93 | 61 | 7 | 0.98 | 0.90 |
| 24 | 8 | 1.00 | 1.00 | 62 | 7 | 1.00 | 0.99 |
| 25 | 11 | 1.00 | 1.00 | 63 | 5 | 1.00 | 1.00 |
| 26 | 4 | 1.00 | 1.00 | 64 | 4 | 1.00 | 1.00 |
| 27 | 9 | 1.00 | 1.00 | 65 | 9 | 1.00 | 0.94 |
| 28 | 7 | 0.97 | 0.99 | 66 | 15 | 1.00 | 0.97 |
| 29 | 8 | 1.00 | 0.97 | 67 | 8 | 0.91 | 0.90 |

Table 5.2: Evaluation of Instance-Based clustered Composition Contd.

| **Comp. #** | **No. of** | $e_{tc}$ | $e_{tc}$ | **Comp. #** | **No. of** | $e_{tc}$ | $e_{tc}$ |
|---|---|---|---|---|---|---|---|
| | **Segs** | **Broadcast** | **Clustered** | | **Segs** | **Broadcast** | **Clustered** |
| 30 | 9 | 0.80 | 0.68 | 68 | 11 | 1.00 | 1.00 |
| 31 | 12 | 0.93 | 0.76 | 69 | 8 | 1.00 | 0.96 |
| 32 | 5 | 0.96 | 0.97 | 70 | 7 | 1.00 | 0.94 |
| 33 | 4 | 1.00 | 1.00 | 71 | 5 | 0.96 | 0.97 |
| 34 | 9 | 0.99 | 1.00 | 72 | 5 | 0.99 | 0.96 |
| 35 | 7 | 0.98 | 0.99 | 73 | 6 | 1.00 | 1.00 |
| 36 | 6 | 1.00 | 0.91 | 74 | 8 | 1.00 | 1.00 |
| 37 | 7 | 0.86 | 0.93 | 75 | 6 | 1.00 | 1.00 |
| 38 | 6 | 0.96 | 0.87 | 76 | 7 | 1.00 | 1.00 |

## 5.4 Thematic Composition

By using the data from the first two compositions in Table 5.4 we composed two news items with a constant similarity threshold value $\tau = 1$ and different values of dissimilarity threshold $\lambda$ to study the thematic composition technique. The results for these two news items are shown in Tables 5.5 and 5.6. Note that identical results were obtained within the values ranges specified in the table for $\lambda$.

The results show that as the value of $\lambda$ increases the thematic continuity increases and the information $(In)$ value decreases. The number of segments in a composition, temporal continuity, and period span covered, do not show distinct patterns. This is because different values of $\lambda$ lead to compositions following different threads in the storyline. For lower values of $\lambda$, all threads

Table 5.3: Evaluation of Instance-Based Thematic Composition

| **Comp.** | **No. of** | **Not-Clustered** | **Clustered** | **Comp.** | **Actual** | **Not-Clustered** | **Clustered** |
|---|---|---|---|---|---|---|---|
| **#** | **Segs** | $e_{thc}$ | $e_{thc}$ | **#** | **Segs** | $e_{thc}$ | $e_{thc}$ |
| 1 | 9 | 0.97 | 0.96 | 2 | 9 | 0.93 | 0.52 |
| 3 | 8 | 0.95 | 0.95 | 4 | 7 | 1.00 | 1.00 |
| 5 | 7 | 0.95 | 0.91 | 6 | 7 | 0.85 | 0.76 |
| 7 | 10 | 0.87 | 0.87 | 8 | 6 | 1.00 | 0.98 |
| 9 | 10 | 1.00 | 0.96 | 10 | 6 | 0.98 | 0.99 |
| 11 | 7 | 1.00 | 1.00 | 12 | 5 | 1.00 | 0.98 |
| 13 | 6 | 0.90 | 0.93 | 14 | 8 | 0.99 | 1.00 |
| 15 | 8 | 0.98 | 0.99 | 16 | 7 | 1.00 | 1.00 |
| 17 | 4 | 1.00 | 1.00 | 18 | 7 | 0.98 | 0.99 |
| 19 | 14 | 1.00 | 1.00 | 20 | 12 | 1.00 | 1.00 |
| 21 | 4 | 1.00 | 1.00 | 22 | 12 | 1.00 | 1.00 |
| 23 | 11 | 0.99 | 0.98 | 24 | 8 | 1.00 | 1.00 |
| 25 | 11 | 1.00 | 0.99 | 26 | 4 | 1.00 | 1.00 |
| 27 | 9 | 1.00 | 1.00 | 28 | 7 | 1.00 | 0.99 |
| 29 | 8 | 1.00 | 1.00 | 30 | 9 | 0.81 | 0.78 |
| 31 | 12 | 0.91 | 0.89 | 32 | 5 | 0.97 | 0.97 |
| 33 | 4 | 1.00 | 1.00 | 34 | 9 | 1.00 | 1.00 |
| 35 | 7 | 1.00 | 1.00 | 36 | 6 | 1.00 | 0.99 |
| 37 | 4 | 0.93 | 0.87 | 38 | 6 | 0.97 | 0.88 |
| 39 | 4 | 1.00 | 1.00 | 40 | 7 | 1.00 | 1.00 |
| 41 | 10 | 1.00 | 1.00 | 42 | 7 | 1.00 | 1.00 |
| 43 | 5 | 1.00 | 1.00 | 44 | 5 | 0.99 | 0.99 |
| 45 | 7 | 1.00 | 1.00 | 46 | 4 | 1.00 | 1.00 |
| 47 | 5 | 0.97 | 0.97 | 48 | 5 | 1.00 | 1.00 |
| 49 | 10 | 0.99 | 0.99 | 50 | 8 | 0.99 | 1.00 |

Table 5.4: Evaluation of Period-Based Temporal Ordering

| **Composition** | **No. of Segs** | **Span** | $e_{thc}$ | $e_{cp}$ |
|---:|---:|:---:|:---:|:---:|
| 1 | 18 | 01/29/1998, 18:38:43 - 01/30/1998, 09:09:21 | 0.85 | 0.91 |
| 2 | 91 | 01/21/1998, 18:30:00 - 02/04/1998, 22:13:40 | 0.92 | 0.96 |
| 3 | 79 | 01/20/1998, 20:13:20 - 02/04/1998, 22:17:00 | 0.72 | 0.96 |
| 4 | 31 | 01/19/1998, 23:04:01 - 01/25/1998, 18:06:15 | 1.00 | 0.96 |
| 5 | 19 | 01/29/1998, 12:00:00 - 02/02/1998, 22:05:16 | 1.00 | 0.94 |
| 6 | 18 | 01/29/1998, 18:38:43 - 02/05/1998, 22:39:08 | 0.89 | 0.95 |
| 7 | 17 | 02/07/1998, 22:20:18 - 02/18/1998, 20:29:56 | 0.99 | 0.98 |
| 8 | 36 | 01/29/1998, 18:38:43 - 02/18/1998, 20:29:56 | 0.99 | 0.97 |
| 9 | 22 | 01/21/1998, 18:30:00 - 01/23/1998, 18:32:06 | 0.85 | 0.94 |
| 10 | 32 | 01/21/1998, 20:00:09 - 01/23/1998, 18:32:06 | 0.88 | 0.96 |
| 11 | 22 | 01/24/1998, 18:30:16 - 01/28/1998, 09:15:23 | 1.00 | 0.98 |
| 12 | 24 | 01/28/1998, 18:33:25 - 02/01/1998, 08:00:21 | 1.00 | 0.95 |
| 13 | 9 | 01/23/1998, 18:55:12 - 01/26/1998, 09:22:24 | 1.00 | 0.97 |
| 14 | 8 | 02/04/1998, 22:22:20 - 02/05/1998, 22:12:18 | 0.98 | 0.95 |
| 15 | 29 | 01/29/1998, 18:49:22 - 02/03/1998, 22:03:51 | 1.00 | 0.93 |
| 16 | 6 | 11/16/1997, 13:12:00 - 01/28/1998, 23:07:41 | 0.99 | 1.00 |
| 17 | 20 | 01/20/1998, 20:24:54 - 02/09/1998, 09:19:19 | 1.00 | 0.91 |
| 18 | 13 | 01/19/1998, 23:04:01 - 01/21/1998, 18:39:05 | 1.00 | 0.93 |
| 19 | 15 | 01/23/1998, 18:49:30 - 01/25/1998, 18:06:15 | 0.97 | 0.95 |
| 20 | 8 | 01/20/1998, 20:00:37 - 01/20/1998, 20:05:18 | 1.00 | 1.00 |
| 21 | 26 | 01/24/1998, 08:20:35 - 01/28/1998, 09:03:53 | 1.00 | 0.97 |
| 22 | 24 | 01/26/1998, 18:48:09 - 01/28/1998, 09:03:53 | 1.00 | 0.97 |
| 23 | 15 | 01/20/1998, 20:13:20 - 01/21/1998, 20:24:52 | 0.99 | 0.92 |
| 24 | 37 | 01/24/1998, 18:32:03 - 01/30/1998, 18:34:43 | 0.97 | 0.97 |
| 25 | 35 | 01/27/1998, 18:39:50 - 02/02/1998, 09:01:18 | 0.99 | 0.96 |
| 26 | 30 | 02/02/1998, 22:11:11 - 02/07/1998, 22:07:07 | 1.00 | 0.99 |
| 27 | 31 | 01/30/1998, 22:10:12 - 02/07/1998, 22:07:07 | 0.99 | 0.97 |
| 28 | 27 | 02/07/1998, 22:05:03 - 02/13/1998, 20:01:20 | 1.00 | 1.00 |
| 29 | 26 | 02/09/1998, 09:01:47 - 02/15/1998, 21:05:40 | 1.00 | 1.00 |
| 30 | 151 | 01/20/1998, 20:13:20 - 02/15/1998, 21:05:40 | 0.79 | 0.98 |

are included in a composition with low thematic continuity. The automatic composition has relatively high thematic continuity as compared to the reference broadcast news.

Table 5.5: Evaluation of Thematic Continuity: Composition 1

| **Comp. #** | **No. of Segs** | $\lambda$ | $In$ | $e_{thc}$ | $e_{cp}$ | $e_{tc}$ | $e_{ps}$ |
|---:|---:|---:|---:|---:|---:|---:|---:|
| 1 | 18 | 0.1 - 0.42 | 1.0 | 0.86 | 0.91 | 1.00 | 1.00 |
| 2 | 15 | 0.43 | 0.82 | 0.93 | 0.94 | 1.00 | 1.00 |
| 3 | 14 | 0.44 - 0.46 | 0.77 | 0.89 | 0.93 | 1.00 | 1.00 |
| 4 | 4 | 0.47 - 0.5 | 0.19 | 0.90 | 0.96 | 1.00 | 0.0008 |
| 5 | 3 | 0.51 - 0.52 | 0.14 | 0.93 | 0.95 | 1.00 | 0.0007 |
| 6 | 2 | 0.53 - 0.56 | 0.10 | 0.99 | 0.93 | 1.00 | 0.0007 |
| 7 | 2 | 0.57 - 0.58 | 0.09 | 1.00 | 1.00 | 1.00 | 0.0008 |
| 8 | 8 | 0.59 | 0.42 | 1.00 | 0.96 | 1.00 | 0.99 |
| 9 | 5 | 0.6 | 0.25 | 1.00 | 0.95 | 1.00 | 0.99 |
| 10 | 4 | 0.61 | 0.19 | 1.00 | 0.96 | 1.00 | 0.99 |
| 11 | 3 | 0.62 | 0.14 | 1.00 | 0.95 | 1.00 | 0.99 |
| 12 | 1 | 0.63 - 1.00 | 0.05 | NA | 1.00 | NA | 0.00 |

Observing the pattern of a number of segments selected in thematic composition of various storylines (Table 5.7) we see that there are the common concepts among the threads. If the concepts are less common then very few segments are selected.

## 5.5 Thematic Nearness Composition

Thematic nearness composition is studied for the first two compositions of Table 5.4. The results are tabulated in Tables 5.8 and 5.9. The data indicate

Table 5.6: Evaluation of Thematic Continuity: Composition 2

| **Comp. #** | **No. of Segs** | $\lambda$ | $In$ | $e_{thc}$ | $e_{cp}$ | $e_{tc}$ | $e_{ps}$ |
|---|---|---|---|---|---|---|---|
| 1 | 91 | 0.10 - 0.47 | 1.00 | 0.92 | 0.96 | 0.99 | 1.00 |
| 2 | 4 | 0.48 - 0.58 | 0.037 | 1.00 | 0.95 | 1.00 | 0.000056 |
| 3 | 6 | 0.59 - 0.61 | 0.06 | 1.00 | 1.00 | 1.00 | 0.004463 |
| 4 | 3 | 0.62 - 0.63 | 0.02 | 1.00 | 1.00 | 1.00 | 0.000047 |
| 5 | 2 | 0.64 - 0.74 | 0.018 | 1.00 | 1.00 | 1.00 | 0.000047 |
| 6 | 1 | 0.75 - 1.00 | 0.008 | NA | 1.00 | NA | 0.00 |

that thematic continuity is usually not as high as compared to the thematic composition technique, but higher than using temporal ordering alone. It remains within the range of the thematic continuity provided by the reference broadcast news. The number of segments incorporated in a composition is most often higher than achieved with the thematic continuity alone. That is, more threads are covered in the composition, and therefore, more information is covered as well. For these compositions the normalization constant, $A$, is 50.

For both the thematic and thematic nearness composition techniques, if the value of $\lambda$ is very low during composition then the value of the thematic continuity remains within the reference values (Table 3.8). However, with increasing $\lambda$, thematic continuity increases but the value of information falls due to the smaller number of segments incorporated in a composition. As $\lambda$ increases, the period span coverage lacks a pattern due to inclusion of different topic threads. The performance of both thematic and thematic nearness composition techniques is highly dependent on the similarity of

Table 5.7: Evaluation of Thematic Continuity

| **Comp. #** | **Candidate Segments** | **Composed Segments** | $In$ | $e_{thc}$ | $e_{cp}$ | $e_{tc}$ | $e_{ps}$ |
|---:|---:|---:|---:|---:|---:|---:|---:|
| 1 | 18 | 4 | 0.20 | 0.88 | 0.97 | 1.00 | .00086 |
| 2 | 91 | 4 | 0.037 | 1.00 | 0.95 | 1.00 | 0.000056 |
| 3 | 79 | 1 | .01 | NA | 0.75 | NA | NA |
| 4 | 31 | 31 | 1.00 | 0.72 | 0.96 | 1.00 | 1.00 |
| 5 | 19 | 19 | 1.00 | 1.00 | 0.94 | 0.97 | 1.00 |
| 6 | 10 | 10 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 7 | 17 | 16 | 0.94 | 0.94 | 0.99 | 1.00 | 1.00 |
| 8 | 36 | 36 | 1.00 | 0.99 | 0.97 | 1.00 | 1.00 |
| 9 | 22 | 4 | 0.17 | 0.99 | 0.95 | 1.00 | 0.000393 |
| 10 | 32 | 30 | 0.94 | 0.93 | 0.97 | 0.99 | 1.00 |
| 11 | 22 | 21 | 0.95 | 0.99 | 0.99 | 0.99 | 1.00 |
| 12 | 24 | 24 | 1.00 | 1.00 | 0.95 | 1.00 | 1.00 |
| 13 | 9 | 9 | 1.00 | 1.00 | 0.97 | 1.00 | 1.00 |
| 14 | 8 | 8 | 1.00 | 0.98 | 0.85 | 1.00 | 1.00 |
| 15 | 29 | 29 | 1.00 | 1.00 | 0.93 | 1.00 | 1.00 |
| 16 | 6 | 5 | 0.83 | 1.00 | 1.00 | 1.00 | 1.00 |
| 17 | 20 | 20 | 1.00 | 1.00 | 0.91 | 1.00 | 1.00 |
| 18 | 13 | 13 | 1.00 | 1.00 | 0.89 | 1.00 | 1.00 |
| 19 | 15 | 15 | 1.00 | 0.97 | 0.96 | 1.00 | 1.00 |
| 20 | 8 | 8 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 21 | 26 | 26 | 1.00 | 1.00 | 0.97 | 1.00 | 1.00 |
| 22 | 24 | 24 | 1.00 | 1.00 | 1.00 | 0.97 | 1.00 |
| 23 | 15 | 1 | .04 | NA | 0.75 | NA | NA |
| 24 | 37 | 37 | 1.00 | 0.97 | 0.97 | 1.00 | 1.00 |
| 25 | 35 | 35 | 1.00 | 0.99 | 0.92 | 1.00 | 1.00 |
| 26 | 30 | 30 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 |
| 27 | 31 | 31 | 1.00 | 0.99 | 0.97 | 1.00 | 1.00 |
| 28 | 27 | 26 | 0.96 | 0.99 | 1.00 | 1.00 | 1.00 |
| 29 | 26 | 25 | 0.96 | 1.00 | 1.00 | 1.00 | 1.00 |
| 30 | 151 | 1 | 0.00 | NA | 0.75 | NA | NA |

Table 5.8: Evaluation of Thematic Nearness: Composition 1

| **Comp. #** | **No. of Segs** | $\lambda$ | $In$ | $e_{thc}$ | $e_{cp}$ | $e_{tc}$ | $e_{ps}$ |
|---:|---:|---:|---:|---:|---:|---:|---:|
| 1 | 18 | $\leq$ 0.00043 | 1.0 | 0.85 | 0.91 | 1.0 | 1.0 |
| 2 | 8 | 0.00044 - 1.0 | 0.42 | 0.80 | 0.90 | 1.0 | 0.0066 |
| 3 | 4 | 1.1 - 1.2 | 0.19 | 0.87 | 0.96 | 1.0 | 0.000861 |
| 4 | 2 | 1.3 - 1.8 | 0.09 | 0.88 | 1.0 | 1.0 | 0.000268 |
| 5 | 1 | $\geq$ 1.9 | 0.05 | NA | 1.0 | NA | 0.0 |

Table 5.9: Evaluation of Thematic Nearness: Composition 2

| **Comp. #** | **No. of Segs** | $\lambda$ | $In$ | $e_{thc}$ | $e_{cp}$ | $e_{tc}$ | $e_{ps}$ |
|---:|---:|---:|---:|---:|---:|---:|---:|
| 1 | 91 | 0.0 - 0.00016 | 1.0 | 0.92 | 0.96 | 0.99 | 1.0 |
| 2 | 79 | 0.00017 - 0.0002 | 0.86 | 0.91 | 0.96 | 0.99 | 0.60 |
| 3 | 54 | 0.00021 - 0.00022 | 0.58 | 0.88 | 0.97 | 0.99 | 0.33 |
| 4 | 18 | 0.00023 - 0.0044 | 0.18 | 0.87 | 0.97 | 1.0 | 0.0047 |
| 5 | 7 | 0.0044 - 0.15 | 0.068 | 0.90 | 0.97 | 1.0 | 0.00023 |
| 6 | 4 | 0.16 - 0.76 | 0.037 | 0.99 | 0.95 | 1.0 | 0.000056 |
| 7 | 2 | 0.77 - 1.35 | 0.017 | 1.0 | 1.0 | 1.0 | 0.000019 |
| 8 | 1 | $\geq$ 1.35 | 0.008 | NA | 1.0 | NA | 0.0 |

concepts among the candidate segments.

## 5.6 Time-Limited Composition

In this section we consider the quality of compositions sequenced under playout time constraints. The effect on thematic continuity and content progression is shown in Table 4.7. The thematic continuity varies between 0.45 and 1.00 and the content progression varies between 0.5 and 1.00.

Next, we evaluate the quality of period-based time-limited compositions. We also compare the performance of the creation-period time-limited algorithm with the trivial scheme. Both the trivial and period-based breadth-first time-limited composition schemes are evaluated based on Composition 2 of Table 5.9 and a $\lambda = 0.00017$.

### 5.6.1 Trivial Scheme

For this adjustment technique we include all sequential segments that fits into the time constraint. Table 5.10 shows the character of these compositions for a range of composition durations applied to the technique.

### 5.6.2 Creation-Period Time Limited Algorithm

Again using Composition 2 from Table 5.10, results are generated based on the technique and are shown in Table 5.11. Here the values of $TP_i$ are constant at 24 hours. The data indicate that span coverage is usually greater as

Table 5.10: Evaluation of Trivial Temporal Adjustment: Composition 2

| **Comp. #** | **Duration** | **No. of Segs** | $In$ | $e_{tc}$ | $e_{cp}$ | $e_{tc}$ | $e_{ps}$ |
|---:|---:|---:|---:|---:|---:|---:|---:|
| 1 | 3,000 | 79 | 0.86 | 0.91 | 0.96 | 0.99 | 0.60 |
| 2 | 2,000 | 76 | 0.83 | 0.91 | 0.96 | 0.99 | 0.60 |
| 3 | 1,000 | 33 | 0.35 | 0.86 | 0.96 | 0.98 | 0.84 |
| 4 | 500 | 13 | 0.13 | 0.88 | 0.98 | 1.0 | 0.0045 |
| 5 | 250 | 7 | 0.068 | 0.90 | 0.97 | 1.0 | 0.00023 |

compared to the trivial approach. The approach also yields less information, temporal continuity, and thematic continuity, but within the references values. As compared to the trivial technique, the creation-period time limited technique provides information over a greater span.

Table 5.11: Evaluation of Creation-Period Time Limited Algorithm: Composition 2

| **Comp. #** | **Duration (Seconds)** | **No. of Segs** | $In$ | $e_{tc}$ | $e_{cp}$ | $e_{tc}$ | $e_{ps}$ |
|---:|---:|---:|---:|---:|---:|---:|---:|
| 1 | 3,000 | 79 | 0.86 | 0.91 | 0.96 | 0.99 | 1.0 |
| 2 | 2,000 | 64 | 0.70 | 0.91 | 0.97 | 0.99 | 0.99 |
| 3 | 1,000 | 40 | 0.44 | 0.90 | 0.96 | 0.98 | 0.89 |
| 4 | 500 | 19 | 0.19 | 0.93 | 0.94 | 0.96 | 0.81 |
| 5 | 250 | 10 | 0.13 | 0.95 | 0.95 | 0.93 | 0.81 |

## 5.7 Observations and Analysis

In the evaluation of video pieces produced by the proposed techniques, we compare the values obtained as the result of the use of the metrics with that of the reference values established from the broadcast news. We assume that the broadcast news is of the quality that is tolerated by the viewers; therefore, the quality of automatic news production, if within the reference values, should be tolerable by the viewers. We have not conducted any user study to evaluate the quality of the automatic news video. The quality of the automatically produced video can be adjusted. For example, the thematic continuity of the composed news can be varied by adjusting the dissimilarity and similarity thresholds. The values for these thresholds used in the composition and evaluations are not absolute. Depending on the requirements of the viewer the thresholds can be altered as needed. For example, if viewers can tolerate rough thematic transitions in a composition then the dissimilarity threshold can be lowered, or only the temporal ordering can be used.

Similarly, the thresholds used with the metrics for quality evaluation and during compositions can be adjusted to suit the viewer's requirements.

System sensitivity for various features (e.g., information, span covered, theme, and content progression) was evaluated as a series of experiments. In Figures 5.2–5.8 we changed the dissimilarity threshold $\lambda$ (Eq. 4.1) using thematic composition and observed the change in the overall quality of a composition. We varied $\lambda$ between 0.0 and 1.0 and observed the change in quality of composition of composition numbers 1–25 shown in Table 5.7. On average there are 27 segments in each composition and the thick line in each figure depicts the average value for the metrics involved.

As seen from Fig. 5.2 the information contained in a composition changes

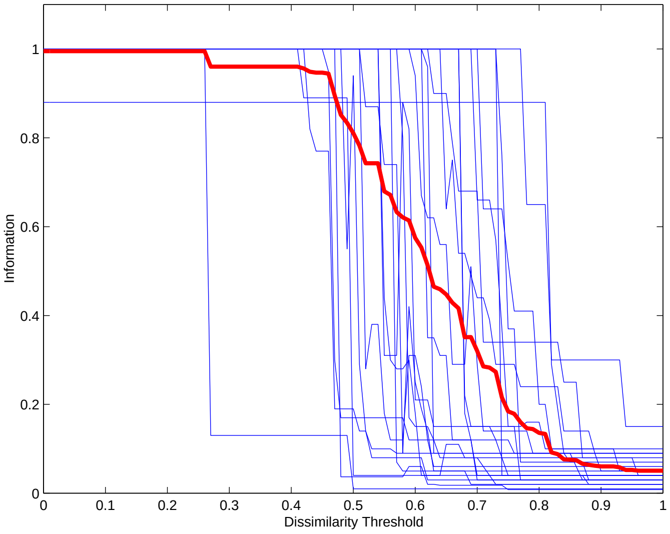


Figure 5.2: Information vs. Dissimilarity Threshold

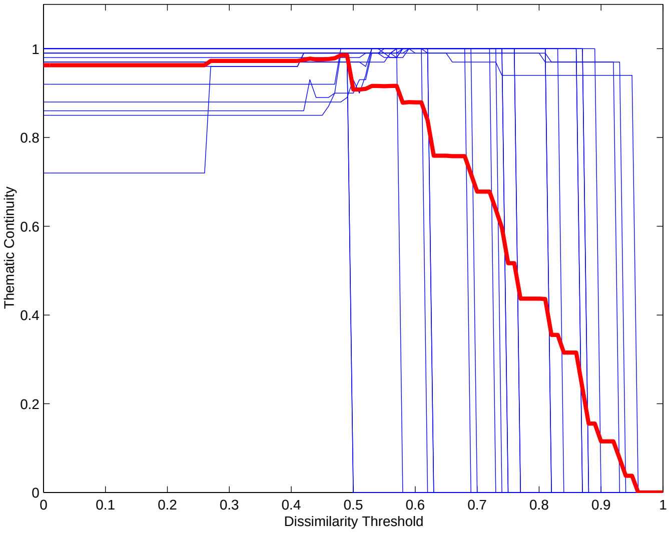


Figure 5.3: Thematic Continuity vs. Dissimilarity Threshold

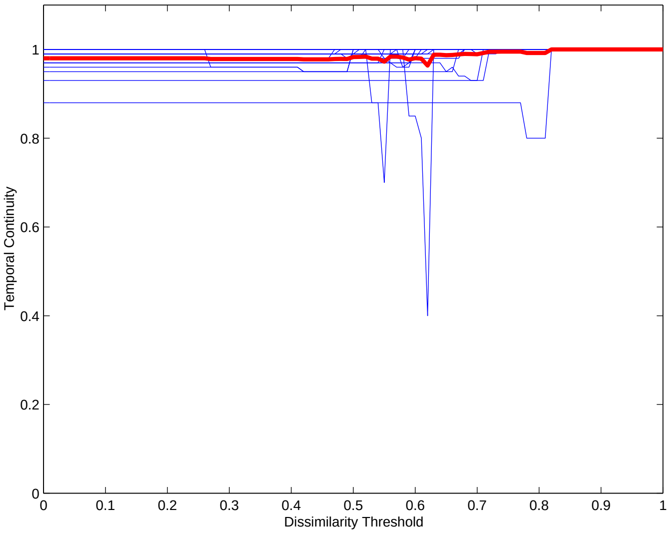


Figure 5.4: Temporal Continuity vs. Dissimilarity Threshold

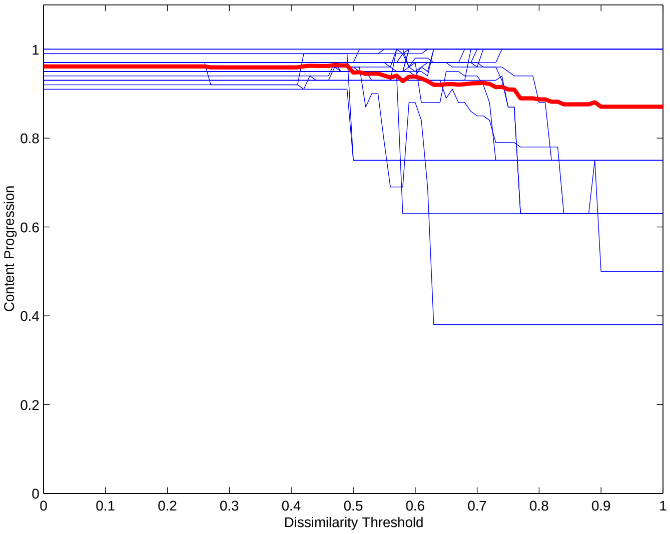


Figure 5.5: Content Progression Quality vs. Dissimilarity Threshold

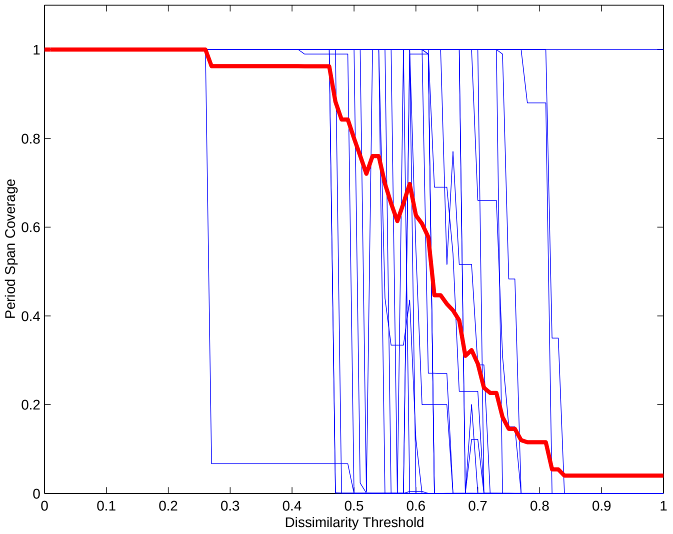


Figure 5.6: Period Span Covered vs. Dissimilarity Threshold

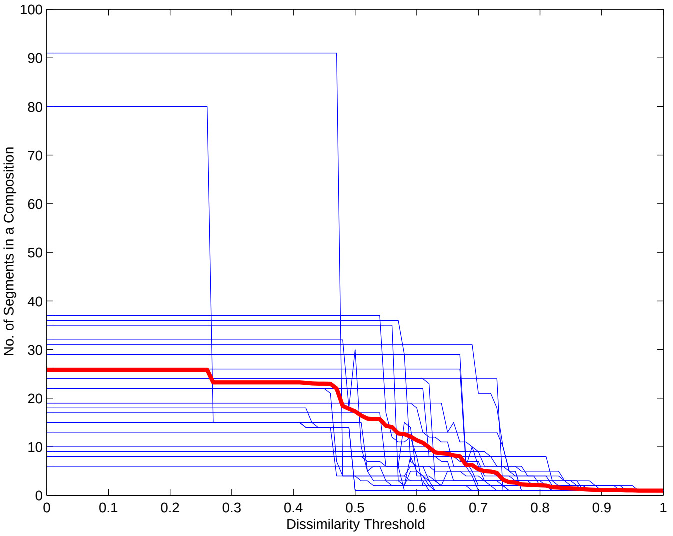


Figure 5.7: Number of Segments in a Composition vs. Dissimilarity Threshold

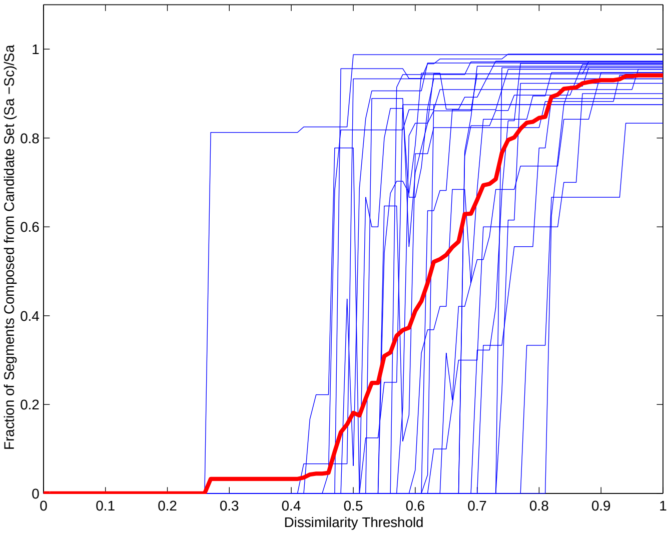


Figure 5.8: Fraction of Segments Composed from Candidate Set vs. Dissimilarity Threshold

and in most of the cases it decreases steadily as $\lambda$ increases. This is due to fewer threads being incorporated in a composition. The pattern of the thematic continuity is similar to the behavior of the quality of information as $\lambda$ varies. In evaluating the thematic continuity (Fig. 5.3) we assumed that the thematic continuity of a composition comprised of one segment is zero.

Content progression of a composition decreases as the $\lambda$ increases. If a large number of segments are present in a composition then the content progression averages to a reasonable value but as the number of segments decrease the change in content progression is more prominent (Figs. 5.7 and 5.8).

In evaluating temporal continuity we assumed that a composition comprised of a single segment has a thematic continuity of one. As observed from Figs. 5.2–5.8 the quality of a composition, on average, is better when $\lambda$ varies between 0.3 and 0.5. Below these values thematic continuity is lower on average. Above these values, thematic continuity, temporal continuity, and content progression show erratic behavior. In addition, the number of segments and information on average decrease appreciably in a composition. Therefore, a dissimilarity threshold between 0.3 and 0.5 is a reasonable operating point for our data set.

For all of the 25 compositions we maintained a similarity threshold ($\tau$) of 1.0. In later observations we found that the variation in the value of $\tau$ does not effect the quality of a composition appreciably; therefore, the selection of $\tau$ equal to 1.0 is not crucial to the quality of a composition.

## 5.8 Summary

The values of thematic continuity evaluated after shuffling segments in instance-based composition is found to be within the range of reference values. The thematic continuity fell below the reference value in one out of 76 compositions evaluated. This shows that in a instance-based composition, random sequencing of segments in a body is possible without degrading its quality.

The values for information and period span metrics for period-based temporal composition are equal to one due to all the segments are incorporated in a composition. For the temporal composition technique we rely on the ordering of segments on the timeline to provide smooth information flow in a composition. By observing the values of thematic continuity we show that the assumption is correct. However, the values are always within the range of reference values, they are not always very high.

For the thematic composition technique, as the value of dissimilarity threshold $\lambda$ increases there is a sudden drop in the number of segments in a composition. This phenomenon is due to the tendency of the technique to follow a smaller number of threads or include segments with similar concepts (i.e., segments that do not have large thematic jumps between them). However, as expected the thematic continuity is higher overall than the output of the temporal composition technique. The results do not show degradation in the temporal continuity as expected, due to the inclusion of only a few segments. The segments do not belong to the threads with large time spans. Also, as individual threads cover small period spans, the spans of the period covered in a composition are small. Consequently, as a result of smaller threads in a composition, information in the composition is low.

For the thematic nearness composition technique, as expected, the values of period span covered and information are higher than the thematic composition technique but not as high as temporal composition technique. On average, the thematic continuity is lower than the thematic continuity of thematic composition but much higher than the temporal composition. The temporal continuity decreases slightly as the period span covered is larger than thematic composition.

The results obtained by analyzing the period-based time-limited technique show that the resulting compositions do not degrade considerably as compared to the original compositions. Furthermore, this technique covers a larger period span as compared with the trivial scheme.

In each of the above evaluations, the automatically composed video pieces where found to be comparable to, or exceed the quality of, the reference broadcast video. This result, based on the defined metrics, both validates our initial assumptions used for creating the composition techniques and demonstrates the viability of automatically composing news video.

# Chapter 6

# Concepts Used in the Design of a News Digital Video Production System

## Synopsis

In this chapter, we present concepts used in design and implementation of various components of a news digital video production system. An ontology is used to establish information and relationships among the information/concepts in a DVPS. A news video data model is used to represent extracted information, and the relationships and the extracted information are stored as metadata. We present observations about semantics in video data, and based on these observations we propose a novel hybrid retrieval technique.

## 6.1 Introduction

A challenging problem in a DVPS is achieving rapid search and retrieval of content from a large video corpus. Because of the computational cost of real-time image-based analysis for searching such large data sets, we pursue techniques based on off-line or semi-automated classification, indexing, and cataloging. We investigate techniques for video concept representation, retrieval, and concept manipulation. In particular, we focus on automatic composition of news stories.

To select and compose video clips in a DVPS, we need to process video data so that they are in clip-queryable form. This is achieved by creating an *ontology* and a *data model*. An ontology consists of a vocabulary (concepts utilized for communicating information to a viewer) needed to extract/annotate information from video clips and establish relationships among the information. A data model is used to represent the extracted information and relationship among the information in a manner that can be used to process user queries and compose video. The concepts/objects that characterize the information contained in video data are called *metadata* [17, 25, 36, 46].

Besides visual, audio transcripts can also be used as a source of metadata because considerable information exists in the audio stream. Based on visual and audio transcript metadata, we propose a novel four-step hybrid approach for retrieval and composition of video newscasts. In the first step, we use conventional techniques to retrieve information from unstructured metadata.

In the second step, unstructured metadata is used to cluster retrieved information into individual news items using a dynamic technique to establish an information cut-off threshold for clustering news items. In the third step, we propose a transitive search technique to increase the recall of the retrieval system. In the final step, we use the union of the different metadata sets to further increase recall performance.

In addition to the composition and customization techniques presented in Chapter 4, we present a grammar and associated production constraints necessary to facilitate automatic video composition in the news domain. The grammar encompasses composition based on content as well as the structure of a newscast. In addition to providing a framework for logical composition of information, the grammar provides constraints for customization of information under bounds on playout duration or content selected by a user.

Next, we discuss the concepts behind the proposed video data model, ontology, annotation techniques, data representation, data selection techniques, and data composition techniques for implementation of a news DVPS.

The symbols used in this chapter are summarized in Table 6.1.

## 6.2 Video Data Ontology and Modeling

For automatic customization of news it is imperative that we understand how a newscast is composed and what elements convey information. The presentation depends greatly on a user's preference of medium and content. For example, a user can seek a topic that comprises text and images; only

Table 6.1: Symbols Used to define Relationships

| **Symbols** | **Descriptions** |
|---|---|
| $R_f$ | A binary relationship on $S$ for transitive search |
| $R_u$ | A binary relationship on $S$ for related segment search |
| $d(a, b)$ | The similarity distance between two sets of keywords |
| $p$ | Production grammar |
| $sa$ | Synthesized attribute of $p$ |
| $NC$ | Total number of news item in the universe |
| $NI$ | A news item consisting of sequenced segments from set $S_c$. |
| $U$ | A set of users |
| $E$ | Edge of a directed graph whose vertices are $NI$ |
| $l(E)$ | Function that maps a user to an edge $E$ |

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 compose, as shown in Table 6.2.

Objects associated with the ontology fall within the categories defined by Rowe et al. [77] (i.e., bibliographical, structural and content based). However, we divide information conveyed by the concept ontology into two categories *structural metadata* and *content metadata* and we define them as:

**Structural Metadata:** Video structure includes media-specific attributes such as recording rate, compression format, and resolution; and cinematographic structure such as frames, shots, sequences, and the spatio-temporal

Table 6.2: News Data 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. |
| 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 [19]. |
| | Voice Over (VO) | When a story uses continuous visuals without showing the speaker. |
| Segment | We divide a newscast item into conceptual segments: | |
| | Headline | Synopsis of the news event. |
| | Introduction | Anchor introduces the story. |
| | Current | Describes the existing situation. |
| | Action footage | Current or wild scenes from the location. |
| | Enclose | Contains the current closing lines. |
| | Reenactment | Accurate scenes of situations that are already past or cannot be filmed [19]. |
| | Complete | A news item which cannot be broken down into previous segments. |
| 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 [19]. |
| | 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. |

characterization of represented objects. These are further decomposed as:

- **Media-specific metadata:** Describing implementation-specific information (e.g., video compression format, playout rate, resolution).

- **Cinematographic structure metadata:** Describing creation-specific information (e.g., title, date recorded, video format, camera motion, lighting conditions, weather; shots, scenes, sequences; object spatio-temporal information).

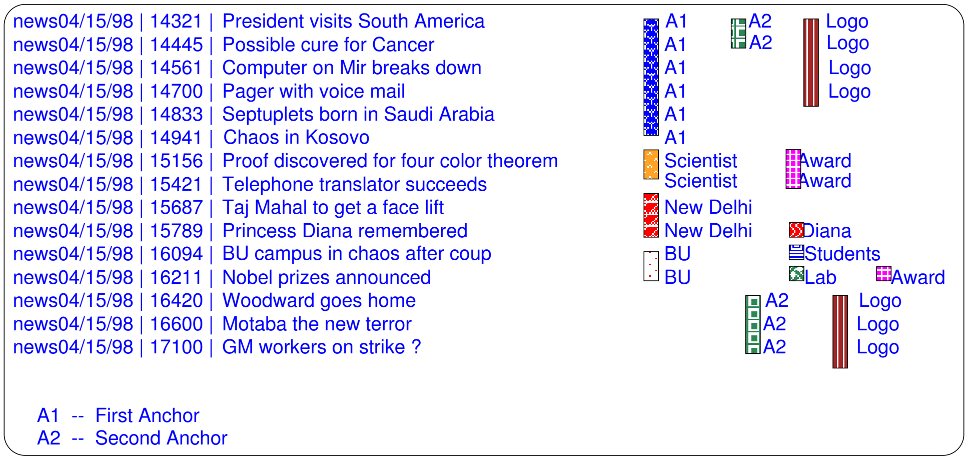
Structural annotations organize linear video sequences as a hierarchy of frames, shots, and scenes [30].

**Content Metadata:** Video content metadata are concerned with objects and meaning in the video stream that appear within or across structural elements. Content metadata are further decomposed as:

- **Tangible objects:** Describing objects that appear as physical entities in the media stream (e.g., a dog, a disc).

- **Conceptual entities:** Describing events, actions, abstract objects, context, and concepts appearing in or resulting from the media stream (e.g., running, catching, tired, master).

A suitable video data model is required to represent object and relationships among them. Two types of techniques are used to model video and associated metadata, *segmentation* [51, 65, 77, 94] and *stratification* [88]. In segmentation, video is divided in groups of contiguous frames or segments that have a start and a end point and metadata are assigned to individual segments. In this process no contextual (concepts) information among the segments is maintained. However this limitation is overcome by stratification; contextual information is segmented into chunks with a begin and end frame (Fig. 6.1).



news04/15/98 | 14321 | President visits South America
news04/15/98 | 14445 | Possible cure for Cancer
news04/15/98 | 14561 | Computer on Mir breaks down
news04/15/98 | 14700 | Pager with voice mail
news04/15/98 | 14833 | Septuplets born in Saudi Arabia
news04/15/98 | 14941 | Chaos in Kosovo
news04/15/98 | 15156 | Proof discovered for four color theorem
news04/15/98 | 15421 | Telephone translator succeeds
news04/15/98 | 15687 | Taj Mahal to get a face lift
news04/15/98 | 15789 | Princess Diana remembered
news04/15/98 | 16094 | BU campus in chaos after coup
news04/15/98 | 16211 | Nobel prizes announced
news04/15/98 | 16420 | Woodward goes home
news04/15/98 | 16600 | Motaba the new terror
news04/15/98 | 17100 | GM workers on strike ?

A1 -- First Anchor
A2 -- Second Anchor

Figure 6.1: Newscast Video Data Model

In the news video data model we utilize both the existing models (i.e., segmentation and stratification). For maintaining structural continuity (i.e., provide complete information) in a composition and to be able to drop segments in playout constrained time composition we manually segment video data. However, to aid in a search, we maintain contextual information across video data. The newscast information model shown in Fig. 6.2 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.[1] 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. Each news item can consist of *1 to 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.

## 6.3 Information Extraction and Representation

The operation of a video database implies the management of a large quantity of raw video data. The presence of this raw data does not significantly assist in indexing and searching. In contrast, video information assists this process. Although any suitable representation can be used to represent metadata, text is commonly used. The concepts in the ontology are captured as tokens (text) and are both domain-dependent and domain-independent and stored

[1] 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.

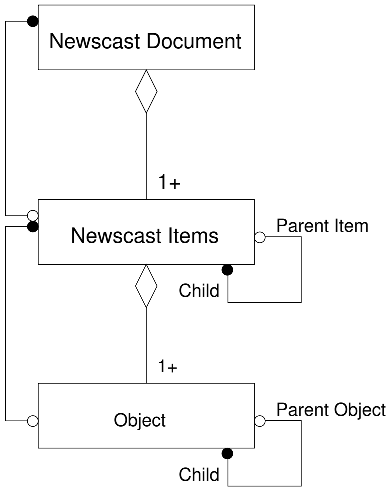


Figure 6.2: Newscast Video Data Model

as metadata (Fig. 6.3).

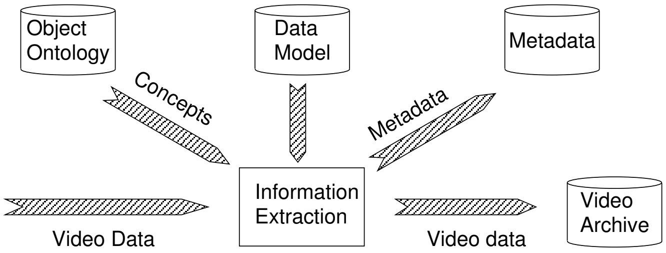


Figure 6.3: Schematic Representation of the Video Information Extraction Process

In addition to extracting concept-based information from the visuals associated with news video data, information from closed caption data can be extracted. Next, we discuss how information can be extracted from various data sets.

### Annotated Metadata

The problem of extracting information becomes one of identifying information contained in the video data and associating it with tokens (metadata). Not surprisingly, humans are quite good at extracting information from video data, whereas it is difficult to get the same performance from an automaton. In the annotation process, a viewer takes notes, albeit biased, of the content of the video stream. During this process, the annotator can be assisted by a computer to provide a more regular representation to capture domain-specific information. For example, a football announcer might use a football-specific metadata schema to capture information about goals scored. In this role, the computer, and the annotation process, provides a consistent and repeatable process for collecting metadata for the application domain.

In Fig. 6.4 we represent the structural objects and the relationship between these objects as a hierarchy and use this representation to store the information in a database. This format is similar to treating a movie as composed of scenes and the scenes composed of shots [7]. The order of the scenes in a movie is identified by the events in the scenes. However, for a newscast the segments are ordered according to their type and creation time under the assumption that all the segments belong to the same event.

The content items within the segments (e.g., Wild Scene) are also treated as objects (e.g., "entity," "location," "category," and "graphics"). An object can be composed of other objects, thus forming a hierarchy of object types. An event, which we treat as synonymous to a news item, forms the root of

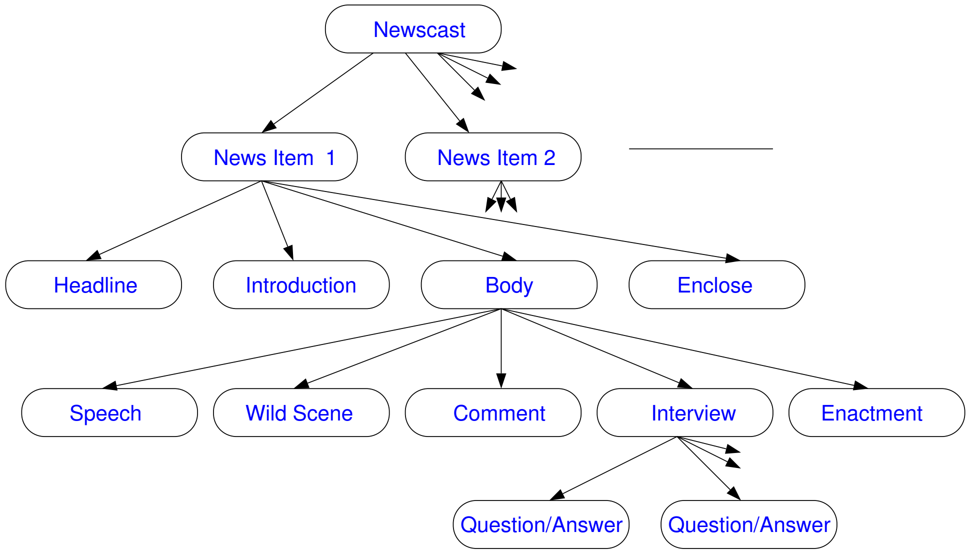


Figure 6.4: Structural Representation for Newscast Composition

an object hierarchy for the news item. Thus, Figs. 6.4 and 6.5 represent the hierarchy of information stored in the metadatabase.

In this hierarchy, Headline, Introduction, Enclose, Speech, Wild Scene, QA, Comment, and Enactment are the *leaves* of the object type tree. Each object is represented by a set of attributes: <object-id, type, name, metatype, medium, popularity, date of creation, time of creation, origin, video-filename, start-frame, end-frame, compression format, playout rate>. The cinematographic attributes "compression format" and "playout rate" are maintained for playout as are the attributes of "video-filename," "start-frame," and "end-frame." Metatype qualifies the type (e.g., an entity-type can be a "person" and its metatype can be "president"). Metatypes are stored so that queries like "give me the reaction of the President" can be satisfied. Headline, In-

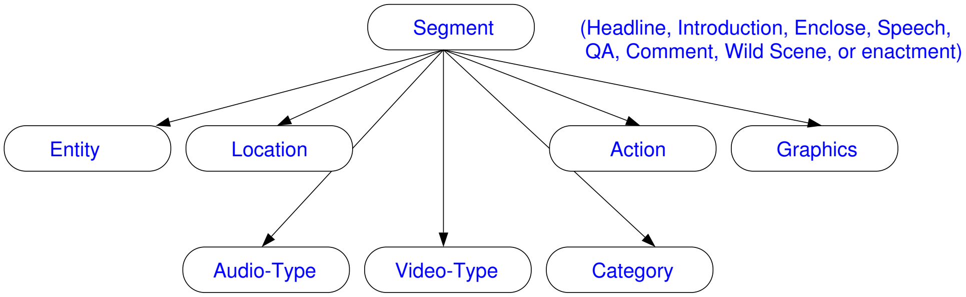

Segment
(Headline, Introduction, Enclose, Speech,
QA, Comment, Wild Scene, or enactment)
Entity
Location
Action
Graphics
Audio-Type
Video-Type
Category

Figure 6.5: Representation of Concepts in a News item

troduction, Wild Scene, Enactment, and Enclose, are the metatypes for "segment." Speech, Interview, and Comment are the metatypes for "reaction." "Country," "city," and "place" are the metatypes for "location." The information whether an object is associated with audio, video, or both audio and video is maintained in the "medium" attribute. The creation time and date represent when an event was recorded. The objects and the information about their attributes are stored as metadata in the form of a regular expression to support automatic composition.

A query can retrieve a set of new items directly by accessing the content metadata. However, for the process of composition, the broader set of metadata need to be used (Section 7.3).

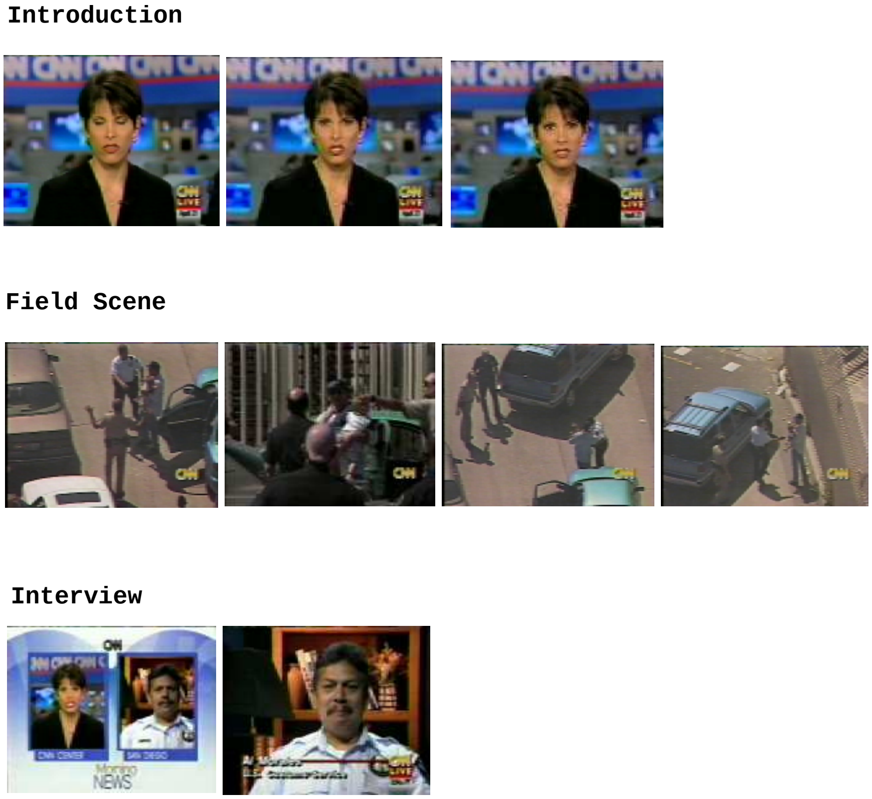
### Unstructured Metadata

In addition to the annotated metadata, transcripts originating from closed-caption data (audio transcripts), when available, are associated with video segments when the segments enter the content universe $S$. These transcripts comprise the unstructured metadata for each segment. Unstructured metadata are used for indexing and forming keyword vectors for each semi-structured metadata segment. Indexing is the process of assigning appropriate terms to a component (document) for its representation.

## 6.4 News Video Data Retrieval

The news video data retrieval techniques presented in this section are an outcome of our observations of generative semantics in the different forms of information associated with news video data. We also studied the common bond among the segments belonging to a single news item.

We observed that synchronized audio and visual data or related video data do not necessarily possess correlated concepts (Fig. 6.6). For example, it is common in broadcast news items that once an event is introduced, in subsequent scenes the critical keywords are alluded to and not specifically mentioned (e.g., Table 6.3, the name Eddie Price is mentioned only in the third scene). However, scenes can share some other keywords, and hence, related by *transitivity*. That is, if scene $a$ is similar to a scene $b$ and the scene $b$ is similar to a scene $c$, then the scenes $a$ and $b$ can be considered similar by transitivity. If a search is made on a person's name, then not all directly related segments are necessarily retrieved. Similarly, related video segments can have different visuals. To rely solely on information contained within transcripts and video data for composition is not prudent. The information tends to vary among the segments related to a news item. Therefore, we



Introduction

Field Scene

Interview

Figure 6.6: Scenes from an Example News Item

Table 6.3: Example Transcripts of Several Scenes

| Introduction | Field Scene | Interview |
|---|---|---|
| A ONE-YEAR-OLD BABY BOY IS SAFE WITH HIS MOTHER THIS MORNING, THE DAY AFTER HIS OWN FATHER USED HIM AS A HOSTAGE. POLICE SAY IT WAS A DESPERATE ATTEMPT TO MAKE IT ACROSS THE MEXICAN BORDER TO AVOID ARREST. CNN'S ANNE MCDERMOTT HAS THE DRAMATIC STORY. | A MAN EMERGED FROM HIS CAR AT THE U.S. MEXICAN BORDER, CARRYING HIS LITTLE SON, AND A KNIFE. WITNESSES SAY HE HELD THE KNIFE TO HIS SON, LATER, TO HIMSELF. AND IT ALL PLAYED OUT ON LIVE TV. OFFICIALS AND POLICE FROM BOTH SIDES OF THE BORDER... | DARYN: JUST IN THE RIGHT PLACE AT RIGHT TIME ESPECIALLY FOR THIS LITTLE BABY. CAN YOU TELL US WHAT YOU WERE SAYING TO THE MAN POLICE IDENTIFIED AS EDDIE PRICE AND WHAT HE WAS SAYING BACK TO YOU? I JUST ASSURED HIM THAT THE BABY WOULD BE OKAY... |

require new techniques to retrieve all the related segments or to improve the recall of the video composition system.

We summarize our observations of video data semantics as follows:

- By utilizing both annotated metadata and closed-caption metadata, precision of the composition system increases. For example, keywords of "Reno, Clinton, fund, raising," if matched against closed-caption metadata, can retrieve information about a place called "Reno" (Nevada). Therefore, annotated metadata can be used to specify that only a person called "Reno" (Janet Reno) should be matched. The results from annotated and closed-captioned searching can be intersected for better precision.

- Recall of a keyword-based search improves if more keywords associated with an event are used. Transcripts provide enriched but unstructured metadata, and can also be used to improve recall. Utilizing transcripts increase the number of keywords in a query; therefore, in some cases precision of the results will be compromised (irrelevant data are retrieved). The transitive search technique is based on this principle (Section 7.3).

- If the relationships among segments of a news event are stored, recall of a system can be increased. For example, if news about "Clinton" is retrieved, then related segment types can be retrieved even if the word "Clinton" is not in them.

As a result of the above observations, we propose a hybrid approach that is based on the union of metadata sets and keyword vector-based clustering as illustrated in Fig. 6.7. We first match a query with unstructured metadata in the universe $S$. The results retrieved (unstructured metadata) are clustered into individual news items.

Next, transitive search is used to augment the clusters. Transitivity on the unstructured data is defined below.

Let $\mathcal{R}_f$ define a binary relationship $f$ on the universal set of video segments $S$ (i.e., $(s_a, s_b) \in R_f \iff s_a$ is *similar* to $s_b$). If similarity distance, defined as $d(s_a, s_b)$ for segments $s_a$ and $s_b$, is greater than an established value then the two segments are considered to be similar. The transitive search satisfies the following property (for all $s_a \in S, s_b \in S, s_c \in S$):

$$(s_a, s_b) \in \mathcal{R}_f \wedge (s_b, s_c) \in \mathcal{R}_f \Rightarrow (s_a, s_c) \in \mathcal{R}_f$$

Therefore, in a transitive search, the results in each cluster are applied

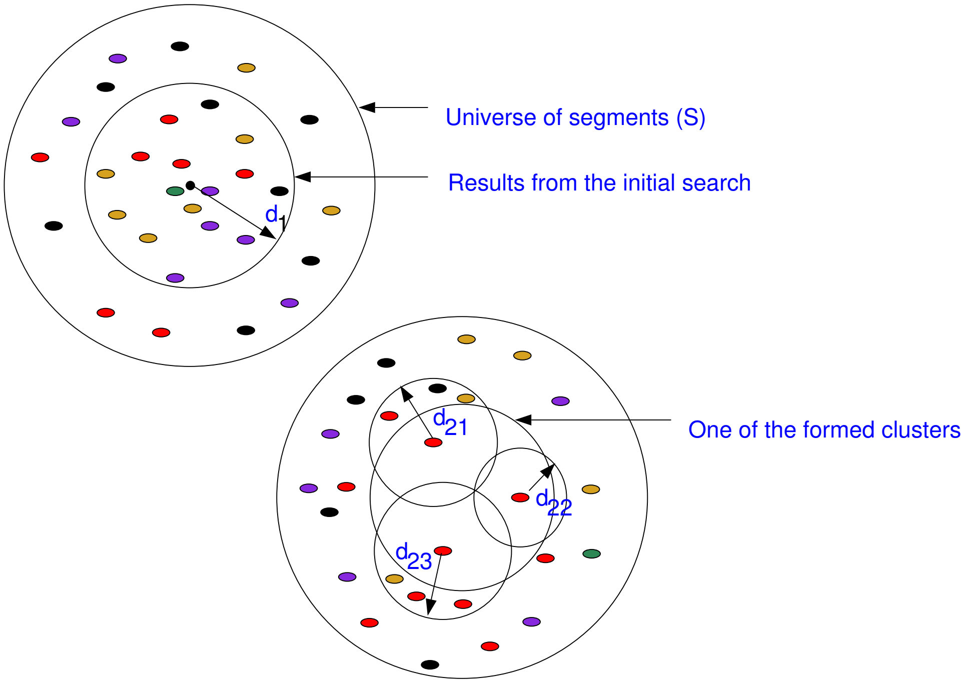


Figure 6.7: Similarity Measure based on the Transitive Search

as a query to retrieve additional unstructured metadata (transcripts) and associated segments, increasing the the recall of the process.

A shortcoming of the aforementioned transitive search is that it may not retrieve all segments related via siblings (related segments). This can be achieved by the following.

Let $\mathcal{R}_u$ define a binary relationship $u$ on the universal set $S$ (i.e., $(s_a, s_b) \in R_u \iff s_a$ and $s_b$ are part of the same news event). As mentioned before, the hierarchical structure of related segments is stored as structural metadata. The final step expands the set of segments as a union operation as follows:

$$S_a \leftarrow S_a \cup \{s_b \mid \exists s_a \in S_a : (s_a, s_b) \in \mathcal{R}_u\},$$

where, $S_a$ is the candidate set of segments used as a pool to generate the final video piece.

The query precision can also be increased by forming the intersection of the keywords from the content and unstructured metadata sets. For example, consider the scenario for composing a news item about Clinton speaking in the White House about the stalemate in the Middle East. From the content metadata, we might be able to retrieve segments of type Speech for this purpose. However, many of the returned segments will not be associated with the topic. In this case an intersection of the query results of the salient keywords applied to the unstructured metadata will give us the desired refinement.

## 6.5 Video Data Composition

Once segments are clustered into respective news items, the segments in the clusters are ordered to form cohesive news items. We use rules specified in grammar of a proposed language to filter and order the segments in a cluster. An EBNF-representation of the language is shown in Table 6.4. The language is defined as a production grammar ($p \rightarrow \gamma$) [9], each symbol $p$ (e.g., <newsitem>) in a production can be interpreted as a node for holding information. The types of information associated with these nodes are defined by the semantic rules of a production.

Production (1) specifies that a newscast is composed of one or more *tours*. For example, a newscast requested by two users can contain the same content but in a different order. Production (2) is a recursively defined rule. The syntactic category or a nonterminal <tour> is defined in terms of itself by right-recursion. A tour is a production of a syntactic category <newsitem> or its recursions.

A tour can be represented as a path in a directed graph. Assume that $NC$ is a set of unique news items that can be formed from the available data and $NC = (NI, E, l)$ is represented as a directed graph. Where $NI$ (news items) are the vertices, $E$ edges that connects vertices, and $l$ is a function from $E$ to a set $U$ of users. The function $l(E)$ assigns the users for whom the edge $E$ can be traversed to compose a news item. In Fig. 6.8, $l(e_2) = (2, 3)$ means that to compose a newscast for users 2 and 3, the edge $e_2$ is traversed in the direction shown. For clarity, $l(E)$ can be written as

Table 6.4: The Proposed Language in EBNF

| | | | | |
|---|---|---|---|---|
| 1. | < newscast > | → {< tour >}$_1^n$ | | |
| 2. | < tour > | → < newsitem > \| < newsitem >< tour > | | |
| 3. | < newsitem > | → < headline > | | |
| 4. | < newsitem > | → [< headline >]{< introduction >}$^1$[< tmp >] | | |
| 5. | < tmp > | → < b-list >< enclose >\|< b-list > | | |
| 6. | < b-list > | → < b-list >< b-list$_2$ >\|< b-list$_2$ > | | |
| 7. | < b-list$_2$ > | → < speech > \| < wild scene > \| < interview > \| < comment > \| < enact | | |
| 8. | < interview > | → < question & answer (qa) > \| < qa >< interview > | | |
| 9. | < headline > | → < shot > | | |
| 10. | | | < headline >.entity-list | := < shot >.entity-list; |
| 11. | | | < headline >.location-list | := < shot >.location-list; |
| 12. | | | < headline >.category-list | := < shot >.category-list; |
| 13. | | | < headline >.event-list | := < shot >.event-list; |
| 14. | | | < headline >.time-list | := < shot >.time-list; |
| 15. | | | < headline >.action-list | := < shot >.action-list; |
| 16. | | | < headline >.graphics-list | := < shot >.graphics-list; |
| 17. | | | < headline >.audio-type-list | := < shot >.audio-type-list; |
| 18. | | | < headline >.video-type-list | := < shot >.video-type-list; |
| 19. | < headline > | → < shot >< headline > | | |
| 20. | | | < headline >.entity-list | := ∪(< shot >.entity-list, < headline > |
| 21. | | | < headline >.location-list | := ∪(< shot >.location-list, < headline |
| 22. | | | < headline >.category-list | := ∪(< shot >.category-list, < headline |
| 23. | | | < headline >.event-list | := ∪(< shot >.event-list, < headline >. |
| 24. | | | < headline >.time-list | := ∪(< shot >.time-list, < headline >.e |
| 25. | | | < headline >.action-list | := ∪(< shot >.action-list, < headline > |
| 26. | | | < headline >.graphics-list | := ∪(< shot >.graphics-list, < headline |
| 27. | | | < headline >.audio-type-list | := ∪(< shot >.audio-type-list, < headli |
| 28. | | | < headline >.video-type-list | := ∪(< shot >.video-type-list, < headli |
| 29. | < introduction > | → < shot > \| < shot >< introduction > | | |

$l(NI_i, NI_j)$, where $NI_i, NI_j$ is an ordered set of vertices which are included in a newscast. The path used to compose a newscast for a user in the graph is simple and elementary (i.e., no news item is visited twice). A news item is presented only once in a newscast for a single user. In Fig. 6.8, path ($e_9$, $e_7$, $e_6$) is traversed to compose a newscast for user 1.

Productions (4) and (5) specify the syntactic category <newsitem> as comprised of <headline>, <introduction>, <b-list>, and <enclose>. A <newsitem> can be composed of only a single headline (see production 3). According to productions (4) and (5) a news item can be produced with a single headline segment, a single introduction segment, a single b-list, and a single enclose segment. An enclose is only present if <b-list> is present. Productions (5), (6), and (7) convey that the syntactic category <b-list> or a body can be composed of any combination of multiple segments belonging to Speech, Wild Scene, Interview, Comment, and Enactment. As mentioned before this kind of composition is valid only if it is based on chronological time. For example, consider a list of segments of type Speech, Interview, Comment, and Wild

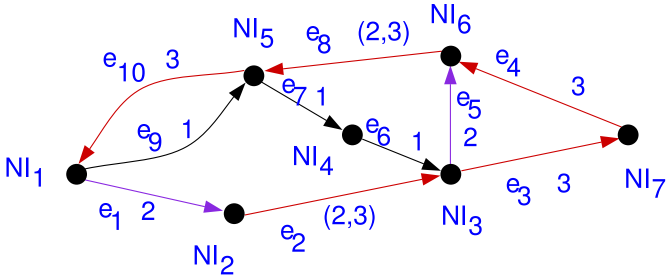


Figure 6.8: Tour Formation from Retrieved News Items

Scene belonging to a body. We show that it is reduced to production (5) as follows:

| | |
|---|---|
| speech, interview, comment, comment, wild scene | |
| b-list$_2$, interview, comment, comment, wild scene | (*production 7*) |
| b-list, interview, comment, comment, wild scene | (*production 6*) |
| b-list, b-list$_2$, comment, comment, wild scene | (*production 7*) |
| b-list, comment, comment, wild scene | (*production 6*) |
| b-list, b-list$_2$, comment, wild scene | (*production 7*) |
| b-list, comment, wild scene | (*production 6*) |
| b-list, b-list$_2$, wild scene | (*production 7*) |
| b-list, wild scene | (*production 6*) |
| b-list, b-list$_2$ | (*production 7*) |
| b-list | (*production 6*) |

Production (8) specifies that a syntactic category <interview> is composed of a "question & answer" or its recursions. The syntactic categories <headline>, <introduction>, <enclose>, <qa>, <speech>, <wild scene>, <comment>, and <enactment> are composed of terminal symbol <shot> or its recursion.

The symbols <headline>, <introduction>, <enclose>, <qa>, <speech>, <wild scene>, <comment>, <enactment> have *synthesized attributes* [9] associated with them. In Table 6.4, entity-list, location-list, category-list, event-list, action-list, graphics-list, audio-type-list, and video-type-list are synthesized attributes of <headline>. Not shown are that these attributes are also associated with other symbols like <introduction>, <enclose>, <qa>, <speech>, <wild-scene>, <comment>, and <enactment>.

An entity-list represents all conceptual (any object part of the commentary, e.g., people) and tangible objects (objects part of a video stream). A location-list consists of all locations shown in the video or conceptual locations, i.e., associations with certain places and countries that are discussed but not part of the visuals (e.g., a news item with discussion on Iraq or shots taken in Baghdad). A category-list consists of the classification of the video data (e.g., accidents, political, sports). An event/action-list represents any happening in a news item (e.g., Clinton's controversy, standoff in Iraq, games at Nagano). A time-list contains the historical time or date of an event or when the event actually took place (e.g., 19 February 1878, phonograph invented by Thomas Edison). A graphics-list represents stills or graphics shown in video (e.g., photographs, maps). An audio-type-list represents the type of audio (i.e., lip-sync, when audio requires tight synchronization with the video), wild-dialogue (dialogue that does not sync with a visible speaker), and voice over (when a story uses continuous visuals without showing the speaker). A video-type-list represents the type of video shots (e.g., close-up shot and wild scene).

Each production grammar $p \rightarrow A_1\,A_2 \ldots A_n$ has an associated set of semantic rules of the form $p.sa := f(A_1.a_1, A_2.a_2, A_3.a_3, ..., A_4.a_n)$, where $sa$ is a synthesized attribute of $p$, $f$ is a function, and $a_1, a_2, ..., a_n$ are the attributes belonging to the symbols of the production grammar. Consider the nodes <headline> → <shot> and <headline> → <shot><headline> in the parse tree. The value of the attribute <headline>.entity-list at this node is defined by:

| Production | Semantic Rule |
|---|---|
| < headline > → < shot > | < headline > .entity-list := < shot > .entity-list; |
| < headline > → < shot >< headline > | < headline > .entity-list := ∪(< shot > .entity-list, < headline > .entity-list); |

Suppose that a headline segment is comprised of three shots. The first shot has three associated entities (*a, b,* and *c*). The second shot has four associated entities (*a, d, e,* and *f*). The last shot has two associated entities (*c* and *g*). Function ∪ performs a union of the two argument lists passed to it. Therefore, the <headline>.entity-list will consist of entities *a, b, c, d, e,* and *f*. Conceptually this semantic rule means that even if an entity is not present in a complete segment it is still assumed to belong to the complete segment. Next, we present some of the examples that depict the mechanism of the proposed grammar.

In this section we demonstrate how the language can be used to compose and customize a newscast. We assume the acquisition of the following data from two sources about "Clinton's visit to Venezuela" (abbreviated to VTV to accommodate Table 6.5).

As previously explained, we store metadata describing the video segments. In Table 6.5 only the structural metadata are shown. For example, $O_{01}$ is an ID of an object/segment that is of type Introduction and is acquired from the CNN. The creation time and date of the segment is "13:00:00" and

Table 6.5: Sample Metadata

| Object ID | Type | Metatype | Name | Source | Creation Time | Creation date |
|---|---|---|---|---|---|---|
| $O_{01}$ | Segment | Introduction | | CNN | 13:00:00 | 06/26/1996 |
| $O_{02}$ | Segment | Wild Scene | | CNN | 13:00:00 | 06/26/1996 |
| $O_{03}$ | Reaction | Speech | | CNN | 13:00:00 | 06/26/1996 |
| $O_{04}$ | Reaction | Comment | | CNN | 13:00:00 | 06/26/1996 |
| $O_{05}$ | Segment | Body | | CNN | 13:00:00 | 06/26/1996 |
| $O_{06}$ | Event | | VTV | CNN | 13:00:00 | 06/26/1996 |
| $O_{07}$ | Segment | Wild Scene | | CBS | 19:00:00 | 06/26/1996 |
| $O_{08}$ | Segment | Wild Scene | | CBS | 19:00:00 | 06/26/1996 |
| $O_{09}$ | Segment | Enclose | | CBS | 19:00:00 | 06/26/1996 |
| $O_{10}$ | Reaction | Interview | | CBS | 19:00:00 | 06/26/1996 |
| $O_{11}$ | Reaction | QA | | CBS | 19:00:00 | 06/26/1996 |
| $O_{12}$ | Reaction | QA | | CBS | 19:00:00 | 06/26/1996 |
| $O_{13}$ | Reaction | QA | | CBS | 19:00:00 | 06/26/1996 |
| $O_{14}$ | Segment | Body | | CBS | 19:00:00 | 06/26/1996 |
| $O_{15}$ | Segment | Introduction | | CBS | 19:00:00 | 06/26/1996 |
| $O_{16}$ | Event | | VTV | CBS | 19:00:00 | 06/26/1996 |

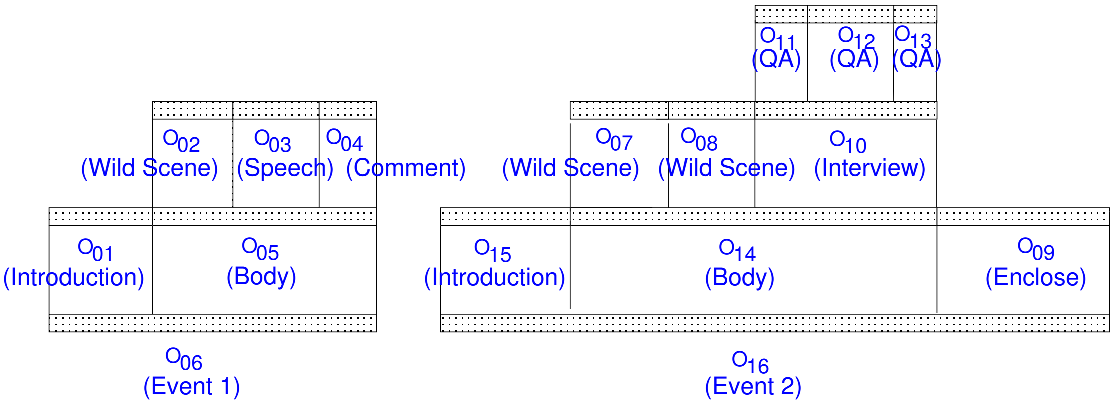


Figure 6.9: Structural Hierarchy for the Content of the Example

"06/26/96" respectively. The hierarchy of the above objects is shown in Fig. 6.9. Object $O_{06}$ is an event and is comprised of two segments/objects $O_{O01}$ and $O_{05}$. Object $O_{05}$ is comprised of three objects $O_{02}$, $O_{03}$, and $O_{04}$. Object $O_{16}$ is another event that is comprised of objects $O_{15}$, $O_{14}$, and $O_{09}$. Object $O_{14}$ is comprised of objects $O_{07}$, $O_{08}$, and $O_{10}$. Finally, object $O_{10}$ is comprised of objects $O_{11}$, $O_{12}$, and $O_{13}$.

With help of queries that are based on the above metadata, we can demonstrate how to form a coherent news item. We also demonstrate how to merge content from various sources, customize content based on a user's temporal constraints, and customize the selection based on content preferences.

**Coherency:** A cohesive news item can be formed by using the language.

*Query 1:* "Compose the news from the most recent material in the system."

In the database the recent objects acquired are from ID $O_{07}$ to $O_{16}$. After finding the objects that are recent (e.g., news less than one hour old), we try to form a coherent composition of the objects for playout. As seen from the object hierarchy (Fig. 6.9), objects $O_{07}$-$O_{15}$ belong to the event ($O_{16}$) "Clinton's visit to Venezuela." We can compose these objects to form a single news item. This is achieved by constraints imposed by the language as follows:

| | |
|---|---|
| $O_{15} \rightarrow O_{14} \rightarrow O_{09}$ | *production 4* |
| $O_{15} \rightarrow O_{07} \rightarrow O_{08} \rightarrow O_{10} \rightarrow O_{09}$ | *productions 5, 6, and 7* |
| $O_{15} \rightarrow O_{07} \rightarrow O_{08} \rightarrow O_{11} \rightarrow O_{12} \rightarrow O_{13} \rightarrow O_{09}$ | *production 8* |

The last row represents the final composition of the news item for playout. It conforms to production rule (4), i.e., there is no segment of type Headline in the news item; and it is composed of a single segment of type Introduction ($O_{15}$), a b-list ($O_{14}$), and a segment of type Enclose ($O_{09}$). The b-list consists of segments $O_{07}$, $O_{08}$, and $O_{10}$. Object $O_{10}$ is further decomposed according to production rule (8). According to production rule (6), a b-list can be composed of segments belonging to the body in any combination. Therefore, segments $O_{07}$, $O_{08}$, and $O_{10}$ can be sequenced in any order.

**Merging:** We can combine content from multiple sources belonging to the same event into a single news item.

*Query 2:* "Retrieve all information on Clinton's visit to Venezuela."

Objects $O_{06}$ and $O_{16}$ are associated with Clinton's trip. All of the subobjects that comprise these two objects can be merged to form a single news item. We require a start, a middle, and an end To form a coherent news item. To maintain temporal continuity and chronology, we include the oldest segment of type Introduction, and the latest segment of type Enclose. Objects belonging to the body are also composed in temporal order (most recent objects shown last). In addition to the language, we impose the additional constraint that all objects in the body appear in chronological order. This constraint is imposed to achieve temporal continuity in presentation. The final composition is as follows:

$$O_{01} \rightarrow O_{02} \rightarrow O_{03} \rightarrow O_{04} \rightarrow O_{07} \rightarrow O_{08} \rightarrow O_{11} \rightarrow O_{12} \rightarrow O_{13} \rightarrow O_{09}$$

Objects $O_{02}$, $O_{03}$, $O_{04}$, $O_{07}$, $O_{08}$, $O_{11}$, $O_{12}$, and $O_{13}$ form the body of the b-list. Object $O_{01}$ is the introduction. Production rule (4) states that an introduction segment is necessary for composition. Object $O_{09}$ is an enclose and is incorporated according to production rule (5). According to our assumptions (Section 4.2) and the constraints imposed by the language, the above composition results is a coherent news item.

**Preferences:** Content-based customization, or "preferences," can be achieved by using the production rules of the language.

*Query 3:* "Retrieve all field shots with information on Clinton's visit to Venezuela."

We gather all the information we have about the event "Clinton's visit to Venezuela" and then apply content-based customization. According to user preferences, only segments of type Wild Scene need to be shown. According to production rule (4) the minimum information to have a coherent news item is an introduction followed by the segments of the type Wild Scene. From the table, objects $O_{02}$, $O_{07}$, and $O_{08}$ belong to type Wild Scene. Using production rules (4), (5), (6), and (7) yields the final composition for the playout as follows:

$$O_{01} \rightarrow O_{02} \rightarrow O_{07} \rightarrow O_{08} \rightarrow O_{09}$$

**Temporal Constraints:** We can achieve time-based customization using the language.

*Query 4:* "Compose the latest news about Clinton's visit to Venezuela for four minutes of playout."

For this type of a query we need to know the playout duration of each clip to produce a news item within the time playout constraint. Assume the timings for the complete playout of objects/segments as shown in Table 6.6.

Here, in addition to using the production rules, we also use the rules for resolving temporal playout constraints. This can be achieved by presenting information from as many views as possible. If an information presentation is from an instance of chronological time, we cluster different views separately (e.g., Wild Scenes, Comments, Interviews). During composition we iterate through the clusters selecting a segment from each cluster (if the playout

Table 6.6: Playout Duration of the Segments

| Object ID | Time (Seconds) |
|---|---|
| $O_{07}$ | 30 |
| $O_{08}$ | 45 |
| $O_{09}$ | 5 |
| $O_{11}$ | 120 |
| $O_{12}$ | 55 |
| $O_{13}$ | 67 |
| $O_{15}$ | 15 |

duration of a segment permits) until the specified duration has been accommodated.

According to the query, we must form a coherent and complete news item within the constraint of $240s$. Event $O_{16}$ and its associated objects have the most recent information; therefore, we initially attempt to compose a news item from these objects with consideration for the duration of each segment. The following objects can be selected to meet the playout constraint of $240s$:

Iteration 1: $O_{15} \rightarrow O_{11} \rightarrow O_{07}$

Iteration 2: $O_{15} \rightarrow O_{11} \rightarrow O_{12} \rightarrow O_{07} \rightarrow O_{09}$

According to the temporal composition criteria, we iterate through the clusters of types QA and Wild Scene. In each iteration we select an object from a cluster until all time is accommodated. If presentation is from a period of chronological time (e.g., from 05/15/1996 to 06/26/1996) we divide the timeline into sub-periods. During composition we iterate through the sub-periods and select a single segment from each sub-period in each iteration.

## 6.6 Summary

In this chapter, we discuss and define types of objects and their attributes that constitute news video information. We define entities as objects (e.g., people, locations, origin, transcripts, graphics, segments, etc.). These types of information/metadata are extracted from video data to process user queries and compose video pieces. Transcripts associated with video data provide additional information and are also used to extract metadata. The extracted information and the relationships among them are represented by a news data model, in which a newscast consists of multiple news items that are, in turn, composed of multiple objects. This data model is used to specify database schema for metadata organization and to process user queries.

We discuss semantics within and across visuals and closed-caption data associated with a news item. We observe that not all segments belonging to the same news item share the same visuals or keywords. Hence, current simple keyword-based retrieval techniques will not retrieve all related data. To overcome this shortcoming (that is, to improve the recall) we propose a novel retrieval technique based on transitive search and the union. The unstructured documents/metadata retrieved as a result of a user query are first clustered/grouped into individual news items. Next, unstructured documents in each cluster are used as queries. As a result of these queries, the retrieved additional unstructured documents are included in the respective clusters. Finally, additional segments are retrieved using sibling relationships among segments. We note that if the results retrieved from use of our proposed four-step hybrid technique are intersected with the results obtained from annotated metadata, the precision of the retrieval system can improve.

We also present a grammar that encompasses the content-based and structure-based constraints to parse a composition. This language is a result of a need for automatic cohesive composition of segments containing desired content. Content alone, though important, cannot be used to create an automatic coherent video piece. Therefore, by incorporating constraints based on both content and structure in the language, it is possible to both, automate and obtain coherence in the news video production process. Using a variety of examples, we demonstrate that the news video production process assisted by the language results in logical composition of newscasts.

# Chapter 7

# Canvass: A News Digital Video Production System

## Synopsis

In this chapter, we discuss the design and implementation of a news digital video production system. We discuss how the information within video data is annotated, what tools are used for this purpose, how metadata are stored in a relational database, and what type of database is used. We present the process of indexing closed-caption/transcripts data and the tools used for indexing. We describe the query processing mechanism for the four-step hybrid data retrieval technique and the quantitative analysis of its performance. We also discuss the implementation of user interface, composition techniques, and video delivery interface.

## 7.1 Introduction

To analyze the quality of compositions resulting from the proposed composition and customization techniques we implemented a news digital video production system. The architecture of the system implementation is shown in Fig. 7.1. Various technologies have been integrated to develop the DVPS and these will be discussed as part of the system architecture.

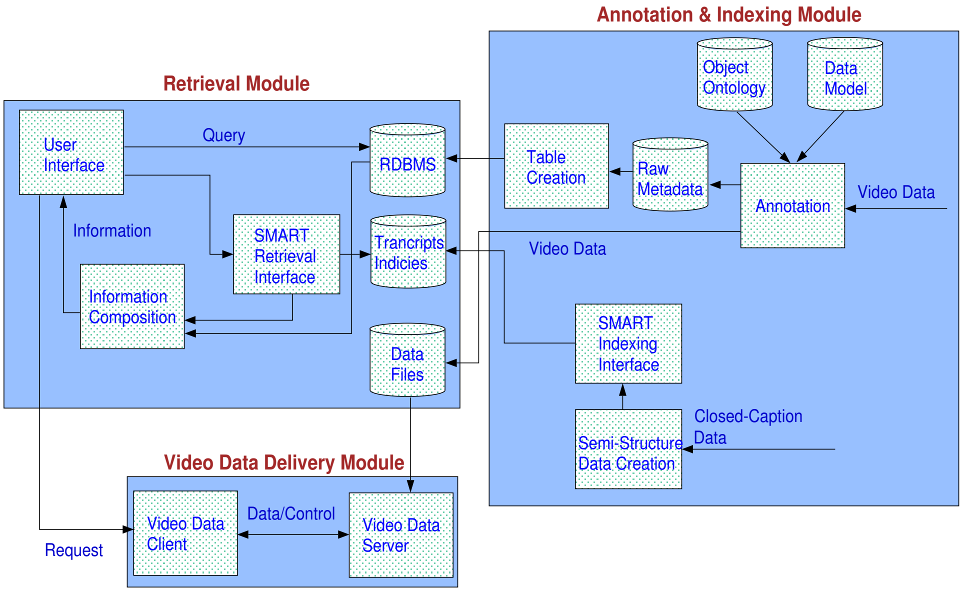


Figure 7.1: Architecture of the Digital Video Production System

The architecture of the news DVPS implemented by us is divided into metadata collection module, retrieval and composition module, and video delivery module. The annotation and indexing module is used to annotate and index video data and transcripts, respectively, and populate a database with the annotated metadata. In the retrieval module we process queries

Table 7.1: Symbols Used to Define the Indexing and the Retrieval Techniques

| **Symbols** | **Descriptions** |
|---|---|
| $K$ | Number of objects in an annotated metadata-based query |
| $M$ | Number of sub-objects of each object type in an annotated metadata-based query |
| $L$ | Number of values selected for each sub-object in an annotated metadata-based query |
| $\mathcal{R}$ | A set of operators used in an annotated metadata-based query |
| $tf_i$ | Frequency of a concept (term) $i$ in unstructured metadata |
| $N_i$ | Number of unstructured metadata components with term $i$ |
| $wt1_i$ | Intermediate weight assigned to a concept $i$ for query match |
| $wt2_i$ | Final weight assigned to a concept $i$ for query match |
| $wt3_i$ | Final weight assigned to a concept $i$ for transitive search |
| $q$ | A query |
| $S_q$ | A set of segments returned as a result of a query |
| $QS$ | A subset of $S_q$ |
| $T_c$ | Cluster cut-off threshold |
| $F$ | Similarity matrix |
| $CL_i$ | A cluster |
| $q(s)$ | A query comprised of unstructured metadata component |
| $s_t$ | A segment retrieved as a result of a query $q(s)$ |
| $S_{q(s)}$ | Set of segments $s_t$ retrieved as a result of a query $q(s)$ |
| $TCL_i$ | An extended cluster $CL_i$ resulting from a transitive search |

using both annotated and transcript metadata and compose the resulting data. The proposed hybrid retrieval techniques and composition techniques are implemented in this module. The video delivery module is used to schedule the playout of actual video segments within a composition. The symbols used in this chapter are summarized in Table 7.1.

## 7.2 Annotation and Indexing

We acquired video data of broadcast news in analog VHS/NTSC format and translated the data into a state suitable for resolving queries to yield candidate sets and composable segments. First, we digitized analog data into MPEG-1 format and stored the data in a repository. Next, the digitized data were used to collect annotated metadata. In parallel, we also recorded closed-caption data associated with the analog video data. The closed-caption data were used to generate unstructured metadata. We discuss this annotation and indexing process in detail.

### 7.2.1 Annotated Metadata

To extract metadata from the digitized video data we used an annotation tool called Vane [22]. This tool is based on the data modeling concepts presented in Chapter 6. The tool supports both segmentation and stratification concepts. This tool is also designed to configure to different concepts in domain-specific ontologies without rewriting the tool itself. This is accomplished by using SGML separating context rules from the information content. The context rules are stored as a document type definition (DTD). The DTD, based on the news video model (Fig. 6.2) and used to configure, Vane is given below:

```
<!--Document Type Definition for Generalized WYSIWYG Example (FULLDOC)-->
```

```
  <!ELEMENT FULLDOC     --    (ABSTRACT?,CATEGORY?,REF*,SEQUENCE*,OBJECT*)>
  <!ELEMENT SEQUENCE    --    (ABSTRACT?,REF*,SCENE*)>
  <!ELEMENT SCENE       --    (ABSTRACT?,SCCATOGR?,REF*,SHOT*,TRANSCR?)>
  <!ELEMENT SHOT        --    (ABSTRACT?,REF*,TRANSCR?)>
  <!ELEMENT OBJECT      --    (REF*,OBJECT*)>
  <!ELEMENT ABSTRACT    --    (#PCDATA & REF*)*>
  <!ELEMENT TRANSCR     --    (#PCDATA)>
  <!ELEMENT REF         -O    EMPTY>
  <!ELEMENT CATEGORY    --    (NEWS)>
  <!ELEMENT SCCATOGR    --    (POLITICS | SPORT | FOREIGN | LOCAL)>
  <!ELEMENT NEWS        -O    EMPTY>
  <!ELEMENT SPORT       -O    EMPTY>
  <!ELEMENT POLITICS    -O    EMPTY>
  <!ELEMENT FOREIGN     -O    EMPTY>
  <!ELEMENT LOCAL       -O    EMPTY>
<!ATTLIST NEWS subcat (morning | mid-day | evening) morning>

<!ATTLIST SPORT subcat (basket | soccer | football | ski | baseball) basket>

<!ATTLIST FULLDOC
  id            CDATA    #IMPLIED
  anchor        CDATA    #CURRENT
  producer      CDATA    #IMPLIED
  location      CDATA    #IMPLIED
  language      CDATA    #IMPLIED
  annotat       CDATA    #CURRENT
  videofile     CDATA    #REQUIRED
  proddate      CDATA    #IMPLIED
  prodtime      CDATA    #IMPLIED
<!ATTLIST SEQUENCE
  id          CDATA    #IMPLIED
  name        CDATA    #REQUIRED
  keyword     CDATA    #CURRENT
  file        CDATA    #CURRENT
<!ATTLIST SCENE
  id           CDATA    #IMPLIED
  name         CDATA    #REQUIRED
  keyword      CDATA    #CURRENT
  populaty     CDATA    #IMPLIED
<!ATTLIST SHOT
  id          CDATA     #IMPLIED
  name        CDATA     #REQUIRED
  keyword     CDATA     #CURRENT
  startf      NUMBER    #REQUIRED
  stopf       NUMBER    #REQUIRED>

<!ATTLIST REF
  target    CDATA    #IMPLIED>
```

```
<!ATTLIST OBJECT
  id          CDATA                          #REQUIRED
  name        CDATA                          #REQUIRED
  type        CDATA                          #IMPLIED
  metatype    CDATA                          #IMPLIED
  creattime   CDATA                          #IMPLIED
  creatdate   CDATA                          #IMPLIED
  medium      CDATA                          #IMPLIED
  origin      CDATA                          #IMPLIED
  populaty    CDATA                          #IMPLIED
  startf      NUMBER                         #REQUIRED
  stopf       NUMBER                         #REQUIRED>
  frate       (30 | 24 | 15)                 30
  mtype       (col |BW)                      col
  mformat     (mpg | cosmo | qt | par | avi)  mpg
```

In the above DTD, the element `FULLDOC` represents the whole video data stream. `SEQUENCE` represents contiguous group of scenes, `SCENE` represents contiguous group of shots, and `SHOT` is a group of frames recorded contiguously. All elements except `REF` have both start and stop tags. For each of the possible contained elements, an occurrence indicator is also expressed. `FULLDOC` can have at most one or possibly no `ABSTRACT - ?` occurrence indicator. As expected, a `FULLDOC` can also have one or more `SEQUENCE`s as represented by "`*`". To support stratification, contextual or content `OBJECTS`s are considered part of `FULLDOC`s and each `OBJECT` can be composed of subobjects. In the same manner, a `SCENE` can have one or more nested `SHOT` elements.

Vane stores raw metadata in a SGML compliant format, hence, we require a translator to populate a database. The translation process includes mapping of metadata to fields in a database schema, populating the data fields, and resolving hypertext references. In the following we describe one translation process that has been constructed to support SQL queries. The translator is called sgml2sql.

Sgml2sql is a conversion tool written to parse the SGML output of the Vane tool and to populate a SQL database. The sgml2sql implementation is modular in nature, built with the premise of supporting enhancements at the production side of the conversion. For example, a change of the database manager affects only the module which interfaces with the database.

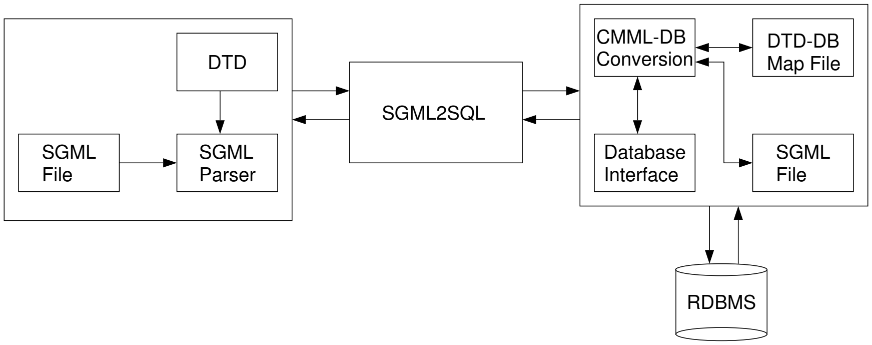


Figure 7.2: Sgml2sql Conversion Tool Architecture

Sgml2sql is written in Perl 5 and uses the DBD/DBI (DataBase Interface) to communicate with the database. Currently we are using the mSQL-DBD package and the mini SQL database. However, the choice of DBMS is not significant for our functionality. Sgml2sql first runs an SGML parser on the SGML file under conversion. The parser checks the SGML file and its associated DTD file for any inconsistencies. If no errors are found at this stage then the tool reads the DTD-to-database-map file, consisting of a mapping between various table attributes to the fields in the database.

The database schema used in the DVPS is shown in Fig. 7.3. 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 a random seek and playout of a 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. Sample annotated metadata from record type `Object` is shown in Table 7.3.

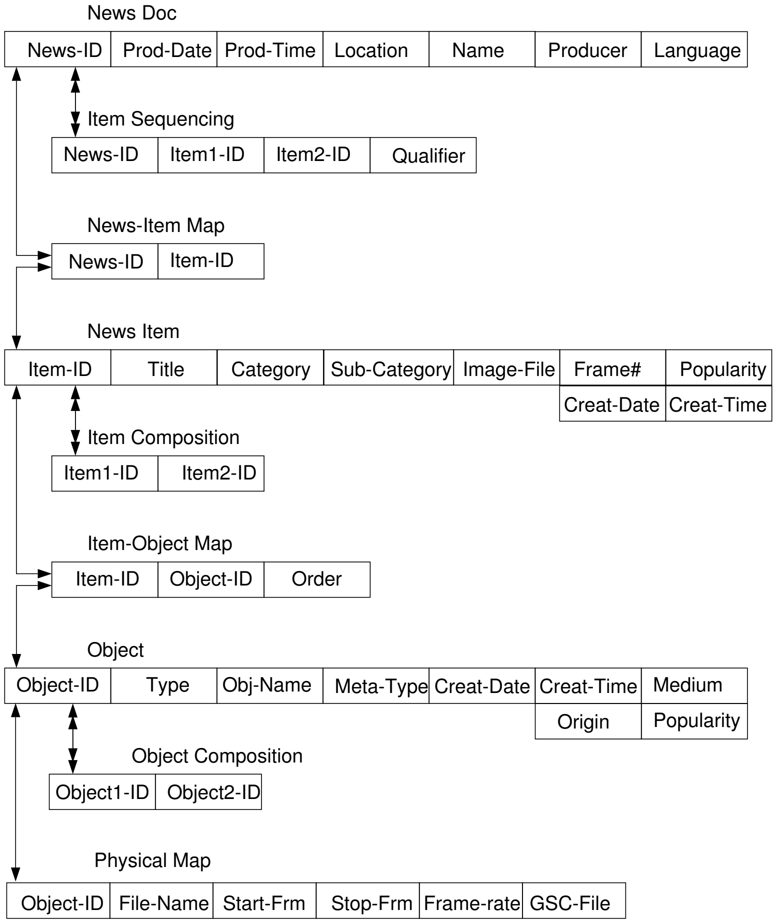


Figure 7.3: Newscast Application-Specific Network Database Schema

An example of the mapping between an DTD and database schema in Fig. 7.3 for the news database is shown in the Table 7.2.

We use a relational database, called miniSQL [58], to store the annotated metadata. Example annotated metadata stored in a relational database (RDB) is shown in Table 7.3.

Table 7.2: Map Between SGML and DB

| SGML Attribute | DB Field |
|---|---|
| fulldoc.id | news.news_id |
| fulldoc.newsid | news.title_id |
| fulldoc.name | news.title |
| fulldoc.producer | news.producer |
| fulldoc.location | news.location |
| fulldoc.language | news.language |
| fulldoc.proddate | news.prod-date |
| fulldoc.prodtime | news.prod-time |
| (fulldoc).sequence | IGNORE |
| (fulldoc).shot | IGNORE |
| (fulldoc).object | new object table entry |
| sequence.(except SCENE) | IGNORE |
| scene | new item entry, new news_item entry |
| scene.id | item.item_id |
| scene.name | item.title |
| scene.keyword | item.keywords |
| scene.imgfile | item.image_file |
| scene.frame | item.frame_num |
| scene.time | item.time |
| scene.date | item.date |
| scene.populaty | item.popularity |
| (scene).sscategor | item.category |
| (scene).sscategor.sucat | item.subcategory |
| (scene).ref | new object and item_object entry |
| (scene).abstract | item.abstract |
| (scene).transcr | uniquely-named file |
| (scene).shot | IGNORE |
| shot.* | IGNORE |

Table 7.3: Example Annotated Metadata

| Object_id | Type | Obj_name | Meta_type | Creat_date | Creat_time | Mediu |
|---|---|---|---|---|---|---|
| O0 | location | Studio | place | 06/26/1996 | 18:00:31 | V |
| O1 | entity | Gene Randal | person | 06/26/1996 | 18:00:31 | AV |
| O2 | segment | NULL | intro | 06/26/1996 | 18:00:31 | NULL |
| O3 | entity | Jamie Mcintyre | person | 06/26/1996 | 18:00:46 | AV |
| O4 | location | Pentagon | place | 06/26/1996 | 18:00:46 | V |
| O5 | audio | Jamie Mcintyre | vo | 06/26/1996 | 18:01:02 | A |
| O6 | graphics | Dhahran | map | 06/26/1996 | 18:01:02 | V |
| O7 | entity | Jamie Mcintyre | person | 06/26/1996 | 18:01:23 | V |
| O8 | location | Pentagon | place | 06/26/1996 | 18:01:23 | V |
| O9 | reaction | NULL | qa | 06/26/1996 | 18:01:23 | NULL |
| O10 | entity | Jamie Mcintyre | person | 06/26/1996 | 18:02:45 | AV |
| O11 | location | Pentagon | place | 06/26/1996 | 18:02:45 | V |
| O12 | reaction | NULL | qa | 06/26/1996 | 18:02:45 | NULL |
| O13 | reaction | NULL | interview | 1996/06/26 | 18:01:23 | NULL |
| O14 | entity | Fionulla Sweeney | person | 06/26/1996 | 18:03:30 | AV |
| O15 | location | Studio | place | 06/26/1996 | 18:03:30 | V |
| O16 | graphics | Dhahran | map | 06/26/1996 | 18:03:45 | V |
| O17 | audio | Abdul Abu Khudair | vo | 06/26/1996 | 18:03:45 | A |
| O93 | entity | Abdul Abu Khudair | person | 06/26/1996 | 18:03:45 | A |
| O18 | location | Jedha | city | 06/26/1996 | 18:03:45 | A |
| O19 | reaction | NULL | qa | 06/26/1996 | 18:03:30 | NULL |
| O20 | entity | Fionulla Sweeney | person | 06/26/1996 | 18:04:36 | AV |
| O21 | location | Studio | place | 06/26/1996 | 18:04:36 | V |
| O22 | graphics | Dhahran | map | 06/26/1996 | 18:04:36 | V |
| O23 | audio | Abdul Abu Khudair | vo | 06/26/1996 | 18:04:36 | A |
| O94 | entity | Abdul Abu Khudair | person | 06/26/1996 | 18:04:36 | A |
| O24 | reaction | NULL | qa | 06/26/1996 | 18:04:36 | NULL |

### 7.2.2 Unstructured Metadata

Transcripts originating from closed-caption data (audio transcripts), when available, are associated with video segments when the segments enter the content universe $S$. These transcripts comprise the unstructured metadata for each segment (Table 7.4).

For indexing the unstructured metadata, we use text indexing and retrieval techniques proposed by Salton [79] and implemented in SMART [21]. To improve recall and precision we use two sets of indices [1], each using different keyword/term weighting schemes as follows:

**Initial Segment Weighting:** Initially, a vector comprised of keywords and their frequency (term frequency *tf*) is constructed using the unstructured metadata of each segment without stemming and without common words. The frequency of a term or keyword indicates the importance of that term in the segment. Next, we normalize the *tf* in each vector with segment (document) frequency in which the term appears by using Eq. 7.1:

$$wt_{1_i} = tf_i \times log\left(\frac{N}{N_i}\right)^2, \tag{7.1}$$

where $N$ is the number of segments in the collection and $N_i$ represents the number of segments to which term $i$ is assigned. The above normalization technique assigns a relatively higher weight $wt_{1_i}$ to a term that is present in a smaller number of segments with respect to the complete unstructured metadata. Finally, $wt_{1_i}$ is again normalized by the length of the vector

Table 7.4: Sample Unstructured Metadata

| |
|---|
| .idDoc:<br>cnn2.txt/O192<br>.videoFile:<br>d64.mps<br>.textData:<br>Leon: Good evening. We begin tonight with attorney general Janet Reno. She says the call was her and she's ready to take the heat. There will be no independent counsel to look into fund-raising activities of the president, vice president, or former energy secretary Hazel O'Leary. |
| .idDoc:<br>cnn2.txt/O193<br>.videoFile:<br>d65.mps<br>.textData:<br>Justice correspondent Pierre Thomas looks at the long-awaited decision. After months of intense pressure, attorney general Janet Reno has made a series of decisions sure to ignite a new round of political warfare. Regarding fund raising telephone calls by Mr. Clinton at the White House: no independent counsel. On vice president Gore's fund raising calls: no independent counsel. Controversial democratic campaign fund-raiser Johnny Chung has alleged he donated $25,000 to O'Leary's favorite charity in exchange for a meeting between O'Leary and a Chinese business associate. Three calls for an independent counsel. All three rejected. |

(Eq. 7.2). Therefore, the influence of segments with longer vectors or more keywords is limited:

$$wt_{2_i} = \frac{wt_{1_i}}{\sqrt{\sum_{j=0}^{n}(wt_{1_j})^2}}. \tag{7.2}$$

**Cluster and Transitive Weighting:** Here we use word stemming along with stop words to make the search sensitive to variants of the same keyword. In segments belonging to a news item, the same word can be used in multiple forms. Therefore, by stemming a word we achieve a better match between segments belonging to the same news item. For the transitive search and clustering, we use the complete unstructured metadata of a segment as a query, resulting in a large keyword vector because we want only the keywords that have a high frequency to influence the matching process. Therefore, we use a lesser degree of normalization (Eq. 7.3) as compared to the initial segment weighting:

$$wt_{3_i} = tf_i \times log\left(\frac{N}{N_i}\right). \tag{7.3}$$

Table 7.5 shows a comparison of the weighting schemes for the same unstructured metadata. The two concepts “Iraq” and “Iraqi” in the second scheme are treated as the same and hence the concept “Iraq” gets a higher relative weight.

Queries are matched against the metadata (annotated and unstructured) and the query formulation and processing process is discussed in the next section.

Table 7.5: Weight Assignment

| *Doc ID* | *Concept* | *Scheme 1* | *Scheme 2* |
|---|---|---|---|
| 146 | barred | 0.62630 | 4.04180 |
| hline 146 | weapons | 0.15533 | 2.50603 |
| 146 | iraqi | 0.21202 | |
| 146 | u.n | 0.18075 | 2.72990 |
| 146 | continues | 0.31821 | 2.58237 |
| 146 | standoff | 0.36409 | 3.87444 |
| 146 | iraq | 0.13211 | 2.71492 |
| 146 | sights | 0.50471 | 4.04180 |

## 7.3 News Video Data Retrieval

The user/query interface is Web-based and it communicates with the news DVPS using Common Gateway Interface (CGI), a standard for external gateway programs to interface with HTTP servers. The CGI scripts are written in the C language. The interface is used to formulate queries to the DVPS. As discussed in the previous chapter, three types of queries can be processed by the DVPS: annotated metadata-based, unstructured metadata-based, and composite metadata-based.

### Annotated Metadata-Based Query Processing

Initially, when a user accesses the Web interface, the interface is populated with the metadata stored in the relational database. Metadata like story/event titles and associated names of the people and location are displayed. The process of query formulation is reduced to the "point and click" process. A user query is converted into miniSQL compliant format for processing (Fig. 7.4).

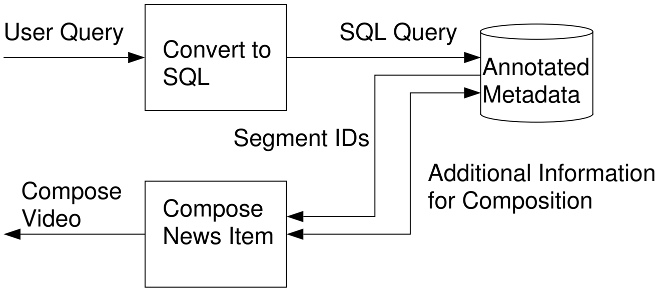


Figure 7.4: Annotated Metadata-Based Query Processing Mechanism

A user query is converted into SQL-compliant format according to the methodology discussed next.

A query predicate $\psi(o)$ is composed of a primitive operation which returns a Boolean result.

$$\psi(o) = o\,\mathcal{R}\,value$$

$o$ represents an object type, $\mathcal{R}$ is a set of operators ($<$, $>$, $=$, $\leq$, $\geq$, $\neq$) which are used to evaluate the tautology of a predicate, and *value* represents any alphanumeric string. For example, city = "Boston." Hence, a general query $Q$ can be expanded as follows:

$$\forall_i : 0 < i \leq N \left| ((\psi(o_{i11}) \prod_{k=2}^{M} \wedge \psi(o_{i1k})) \ \wedge \ (\prod_{j=2}^{L} (\psi(o_{ij1}) \prod_{k=2}^{M} \wedge \psi(o_{ijk}))) \right.$$

Where $N$ represents the number of object types selected for the query (e.g., location and entity) and $L$ represents the number of sub-objects of each object (e.g., sub-objects of location are place, city, and country). In a query, each sub-object can have multiple values (e.g., city can have values "Boston," "San Francisco," and "Srinagar"). $M$ is the number of values selected for each sub-object. An example query is as follows:

> "Retrieve news items about bombing in Dhahran, Saudi Arabia with Clinton and Christopher."

In the above query, a search is made on a city called Dhahran, a country called Saudi Arabia, an event named bombing, and persons (entity) named Clinton and Christopher. Hence, the value of $N$ is 3 (i.e., there are three object types to be searched). The value of $L$ is 2 for the first object (i.e., we have to search for two sub-objects types; city and country). The value of $L$ for other two objects is 1. The value of $M$ for sub-object types city, country, and event is 1. For the last sub-object type (person), the value of $L$ is 2.

$$Q = (city = \text{``}Dahran\text{''})\, AND\, (country = \text{``}Saudia\ Arabia\text{''})\, AND\, (event = \text{``}Bombing\text{''})\, AND\, (person = \text{``}Clinton\text{''}\ AND\, person = \text{``}Christopher\text{''})$$

We store relationships among the segments belonging to a news item as annotated metadata. These metadata are used to form candidate sets.

Therefore, segments that are retrieved as a result of a user query are clustered by finding the relationships among them. These clusters are used to compose news items.

The unstructured metadata-based query is just a string of keywords, and these keywords are matched against the indices created from the unstructured metadata. The proposed query processing technique is a bottom up approach, where the search starts from the unstructured metadata and is discussed in the next section.

### Unstructured Metadata-Based Query Processing

In Section 6.4, we proposed a novel four-step hybrid approach to improve the recall of a video information retrieval system. Here we present the mechanism (Fig. 7.5 ) of processing a query using this approach.

A query enters the system as a string of keywords. These keywords are matched against the indices created from the unstructured metadata. The steps of this process are query matching, clustering the results, retrieval based on the transitive search, and sibling identification. These are described below.

**Query Matching:** This stage involves matching of a user-specified keyword vector with the available unstructured metadata. In this stage we use indices that are obtained as a result of the initial segment weighting discussed in the previous section. As the match is ranked-based, the segments are retrieved in the order of reduced similarity. Therefore, we need to establish a cut-off threshold below which we consider all the segments to be irrelevant

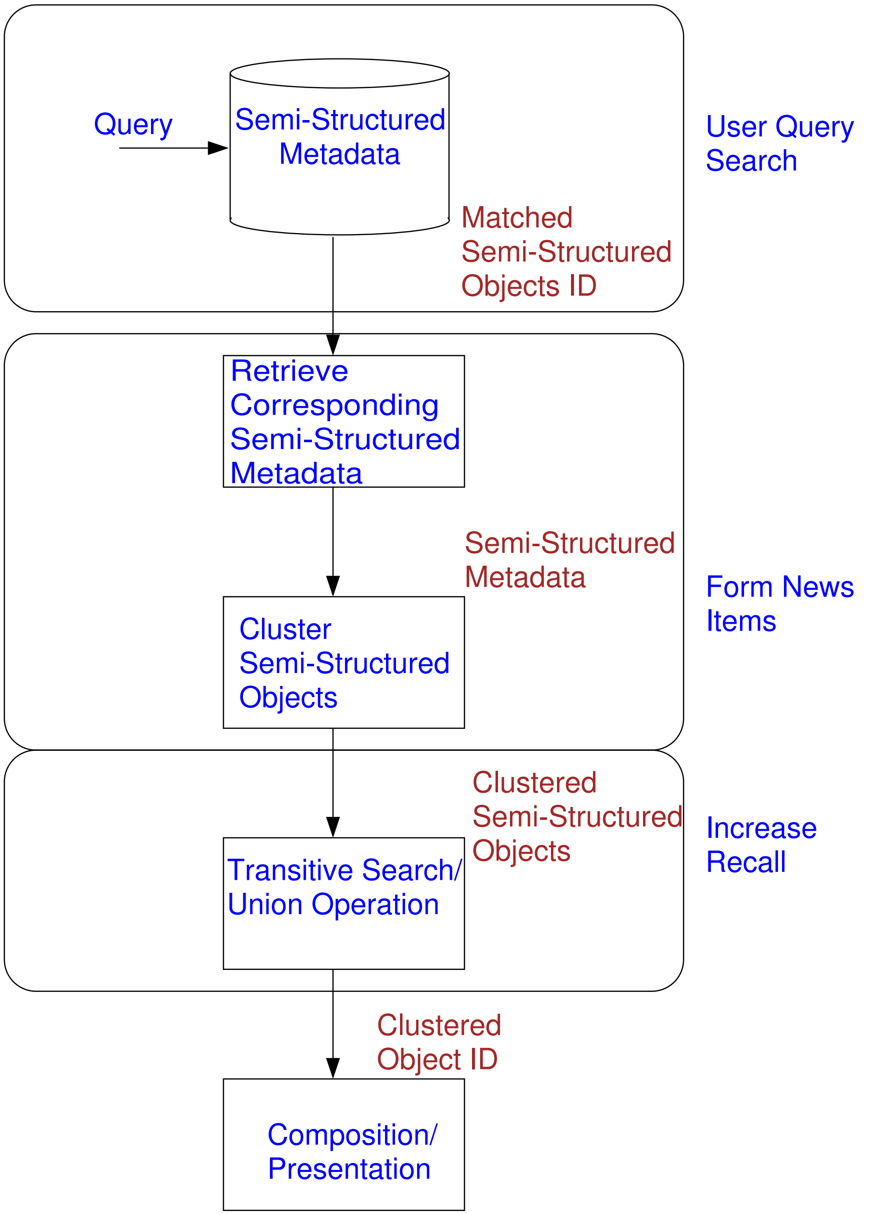

Query
Semi-Structured Metadata
User Query Search
Matched Semi-Structured Objects ID
Retrieve Corresponding Semi-Structured Metadata
Semi-Structured Metadata
Form News Items
Cluster Semi-Structured Objects
Clustered Semi-Structured Objects
Increase Recall
Transitive Search/ Union Operation
Clustered Object ID
Composition/ Presentation

Figure 7.5: Process Diagram for Newscast Video Composition

to the query. Unfortunately, it is difficult to establish an optimum and static query cut-off threshold for all types of queries as the similarity values obtained for each query are different. For example, if we are presented with a query with keywords belonging to multiple news items then the similarity value with an individual object in the corpus will be small. If the query has all keywords relevant to a single news item then the similarity value will be high. Because of this observation, we establish a dynamic query cut-off threshold $(D \times max\{d(s,q)\})$ and we set it as a percentage $D$ of the highest match value $max\{d(s,q)\}$ retrieved in set $S_q$. The resulting set is defined as:

$$QS \leftarrow \{s \in S_q \mid d(s,q) \geq (D \times max\{d(s,q)\})\},$$

where $s$ is the segment retrieved and $d(s,q)$ is the function that measures the similarity distance of segment $s$ returned as a result of a query $q$.

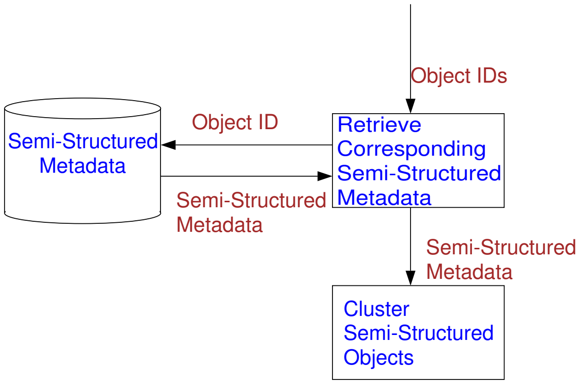


Figure 7.6: Diagram of the Clustering Process

**Results Clustering:** In this stage, we cluster the retrieved segments with each group containing yet more closely related segments (segments belonging to the same event). We use the indices acquired as a result of the transitive scheme (Fig. 7.6). During the clustering process, if the similarity ($d(s_a, s_b)$) of the two segments is within a cluster cut-off threshold $T_c$, then the two segments are considered similar and have a high probability of belonging to the same news event. Likewise, we match all segments and group the segments that have similarity value within the threshold, resulting in a set

$$\{CL_1, CL_2, CL_3, ..., CL_k\},$$

where $CL_i$ are a clusters (sets) each consisting of segments belonging to a single potential news item. An algorithm for forming the clusters is as follows:

For forming disjoint clusters we use a graph-theoristic method [34, 40] that uses minimal spanning tree (MST). The longest edges in the tree are removed producing clusters. In this work, we use a threshold $T_c$ (the edges with length beyond and equal to which are removed) that gives the best clustering performance on the experimental data set. However, if an optimum threshold is to be used, then cluster separation measure proposed by Davies and Bouldin [32] can be used. For creating the MST we use the Prim's algorithm [29] and the depth-first search algorithm to find long edges in the tree. We use the depth-first search due to ease with which the clusters are created. The clusters are formed as follows:

1. If there are $k$ segments in the set $QS$ then first create $k \times k$ similarity

matrix $F = [f_{ij}]$, where

$$f_{ij} = \begin{cases} \frac{1}{d(s_i,s_j)} & \text{if} \quad i \neq j \wedge d(s_i, s_j) > 0 \\ 0 & \text{if} \quad i \neq j \wedge d(s_i, s_j) = 0 \\ 0 & \text{if} \quad i = j \end{cases} \qquad i, j = 1, ..., k$$

2. Use Prim's algorithm for forming MST. The input to the algorithm is the matrix $F$ and output is the tree.

3. Use depth-first traversal through the tree to remove edges greater than the threshold $T_c$. This results in separate clusters $CL_i$ of connected nodes.

**Transitive Retrieval:** We use the transitive search (Fig. 6.7). The transitive search increases the number of segments that can be considered similar. During query matching, the search is constrained to the similarity distance ($d_1$) and segments within this distance are retrieved. During the transitive search we increase the similarity distance of the original query by increasing the keywords in the query so that segments within a larger distance can be considered similar. In the transitive search we use unstructured metadata of each object in every cluster as a query, $q(s)$, and retrieve similar segments. Again, we use item cut-off threshold that is used as a cut-off point for retrieved results and the retained segments are included in the respective cluster.

A news item is made up of a number of segments. Not all segments contain equal level of information. Therefore, a news item is difficult to retrieve from only a few keywords. To retrieve segments that do not match the initial query but belong to same news item we use the complete unstructured metadata for each segment as a query. Related segments have other mutually common keywords that can be used for matching. Therefore, the third stage increases the recall of the initial query by using a transitive search operation.

The transitive cut-off threshold ($T \times max\{d(s_t, q(s))\}$) is set as the percentage ($T$) of the highest similarity value retrieved $max\{d(s_t, q(s))\}$. For example, the distances $d_{21}, d_{22}$, and $d_{23}$ (Fig. 6.7) fall within the transitive cut-off thresholds of respective segments.

Consider a cluster $CL_i = \{s_1, s_2, s_3, ..., s_N\}$ formed in the results clustering step. The extended cluster resulting from the transitive search can be defined as:

$$TCL_i \leftarrow \bigcup_{\forall s \in CL_i} \left\{ s_t \in S_{q(s)} \mid d(s_t, q(s)) \geq (T \times max\{d(s_t, q(s))\}) \right\},$$

where, $s_t$ is a segment returned as a result of a transitive search of a segment $s \in CL_i$ and $d(s_t, q(s))$ is the function that measures the similarity value of a segment $s_t$ to query $q(s)$.

**Sibling Identification:** To further improve recall we use the structural metadata associated with each news item to retrieve all other related objects (Fig. 7.7). Structural information about each segment in a cluster is annotated; therefore, we have the information about all the other segments that are structurally related to a particular segment. We take the set of segments that are structurally related to a segment in a cluster and perform a union operation with the cluster. Suppose $TC_i = \{s_1, s_2, s_3, ...., s_N\}$ is one of the clusters resulting from the third step, then the final set can be defined as:

$$SC_i = \bigcup_{s \in TC_i} R(s).$$

Here $R(s)$ is a set of segments related to a segments $s$. Likewise, the union operation can be performed on the remaining clusters.

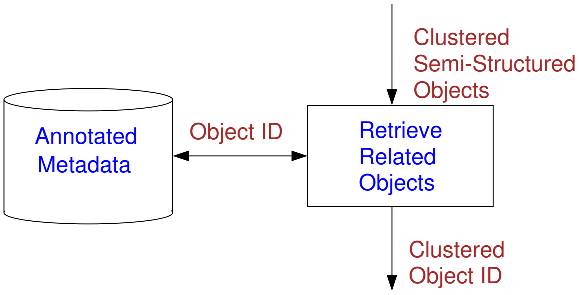


Figure 7.7: Process Diagram for Retrieving Related Objects

Using the four-step hybrid approach we are able to increase the recall of the system. Next we discuss the quantitative analysis of the retrieval, clustering, and proposed transitive search process.

### Analysis of the Proposed Hybrid Technique

We evaluated the performance of our technique based on 10 hours of news video data and their corresponding closed-caption data acquired from the network sources. Our results and analysis of the application of our techniques on this data set are described below.

Because the objective of our technique is to yield a candidate set of video segments suitable for composition, we focus on the inclusion-exclusion metrics of recall and precision for evaluating performance. However, subsequent rank-based refinement on the candidate set yields a composition set that can be ordered for a final video piece.

The data set contains 335 distinct news items obtained from CNN, CBS, and NBC. The news items comprise a universe of 1,731 segments, out of which 537 segments are relevant to the queries executed. The most common stories are about the bombing of an Alabama clinic, Oprah Winfrey's trial, the Italian gondola accident, the UN and Iraq standoff, and the Pope's visit to Cuba. The set of keywords used in various combinations in query formulation is as follows:

race relation cars solar planets falcon reno fund raising
oil boston latin school janet reno kentucky paducah rampage
santiago pope cuba shooting caffeine sid digital genocide
compaq guatemala student chinese adopted girls
israel netanyahu arafat fda irradiation minnesota tobacco trial
oprah beef charged industry fire east cuba beach varadero
pope gay sailor super bowl john elway alabama clinic italy
gondola karla faye tucker death advertisers excavation lebanon
louise woodword ted kaczynski competency

The number of keywords influences the initial retrieval process for each news item used in a query. If more keywords pertain to one news item than the other news items, the system will tend to give higher similarity values to the news items with more keywords. If the query cut-off threshold is high (e.g., 50%), then the news items with weaker similarity matches will not cross the query cut-off threshold (the highest match has a very high value). Therefore, if more than one distinct news item is desired, a query should be composed with an equal number of keywords for each distinct news item. All the distinct retrieved news items will have approximately the same similarity value with the query and will cross the query cut-off threshold.

For the initial experiment we set the query cut-off threshold to 40% of the highest value retrieved as a result of a query, or $0.4 \times max(S_q)$. The transitive cut-off threshold is set to 20% of the highest value retrieved as a result of unstructured metadata query, or $0.2 \times max(S_q(s))$. The results of 29 queries issued to the universe are shown in Fig. 7.8. Here we assume that all the segments matched the query (we consider every retrieved segment a positive match as the segments contain some or all keywords of the query). The clustering threshold $T_c$ was kept at 0.03 and we observed that out of the 29 queries the clustering algorithm did not form exact clusters for 4 of the queries. In all four cases the algorithm could not identify distinct storylines.

Not all the keywords are common among the unstructured metadata of related segments, nor are they always all present in the keywords of a query. Therefore, to enhance the query we use a transitive search with a complete set of unstructured metadata. The probability of a match among related segments increases with the additional keywords; however, this can reduce precision.

As a result of the transitive search, the recall of the system is increased to 48% from 25% (another level of transitive search can increase it further).

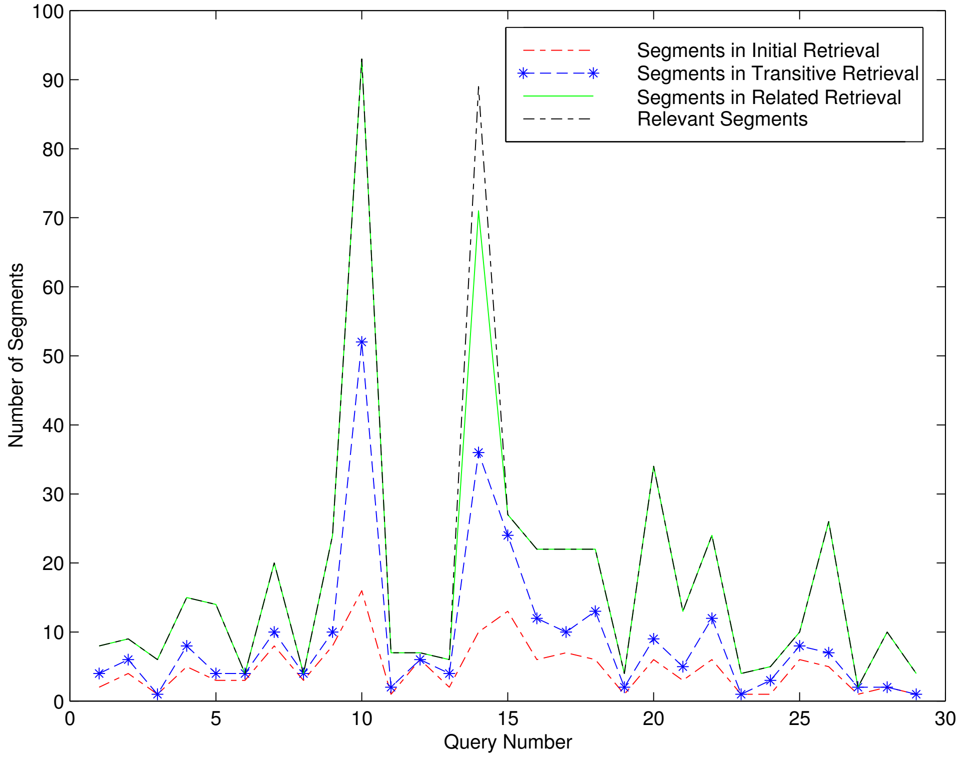


Figure 7.8: Summary of Performance of Different Retrieval Techniques

Table 7.6: System Performance

| Search Technique | Total Segments Retrieved | Relevant Segments Retrieved | Recall | Precision |
|---|---|---|---|---|
| Query Match | 137 | 137 | 25% | 100% |
| Transitive Search | 293 | 262 | 48% | 89% |
| Sibling Identification | 517 | 517 | 96% | 100% |

The precision of the results due to this step is reduced to 89% from 100%.

A cause of such low recall of the initial retrieval and subsequent transitive search is the quality of the unstructured metadata. Often this quality is low due to incomplete or missing sentences and misspelled words (due to real-time human transcription).

Using the structural hierarchy (Section 6.3) we store the relationships among the segments belonging to a news item. Therefore, if this information is exploited we can get an increase in recall without a reduction in precision (as all segments belong to the same news item). In the last step of the query processing we use structural metadata to retrieve these additional segments. As observed from the above results, the recall is then increased to 96%. The remaining data are not identified due to a failure of the prior transitive search.

The results demonstrate that the combination of different retrieval techniques using different sources of metadata can achieve better recall in a news

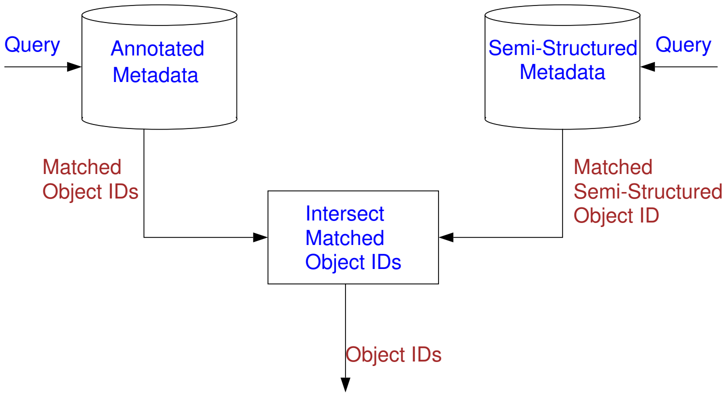


Figure 7.9: Process Diagram for Using Visual Metadata to Increase Precision

video composition system as compared to the use of a single metadata set.

From the results we observed that to emulate news items which encompass multiple foci (i.e., concepts from each are associated with many segments), it becomes difficult to balance the clustering of segments for these foci with our techniques. For example, the query “State of the Union Address” applied to our data set will yield foci for the address and the intern controversy. However, there are many more segments present in the data set for the intern controversy.

### Composite Metadata-Based Query Processing

The results for the annotated metadata-based and unstructured metadata-based search are intersected and only the common segments are retained (Fig. 7.9).

## 7.4 Video Data Delivery

The conceptual compositions or information about a composition is passed to a user and is displayed in the user interface. The video playout is initiated at the user's request by passing the composition information to a video server. The server in turn reads the corresponding video data and sends them to a video client for rendering. The communication between the video server and the client takes place via 100BaseT using TCP/IP protocols. We use the MTV client [62] for video playout.

## 7.5 Summary

In this chapter, we present details about the implementation of a news digital video production system. The analog video data are first converted into MPEG-1 digital format. Using Vane the digital video are annotated. The output of Vane (content information or raw metadata) are stored in a SGML compliant format. To make the raw metadata queryable, it is translated into a relational database specific schema (miniSQL) using the `sgml2sql` tool.

We also decode the closed-caption data associated with video data, and convert them into unstructured metadata. The unstructured metadata are then indexed using SMART and the indices are stored in SMART-compliant files.

Queries are issued using the Web interface, which is implemented using HTML and the Java language. At the time of interface rendering, the annotated metadata are automatically extracted from the RDB and displayed. Queries composed by the user with the "point and click" method are translated into SQL and sent to the relational database (miniSQL) for processing. A user can also enter keywords that are converted into SMART-compliant query format for processing. A user can also simultaneously query both annotated and unstructured metadata.

In the annotated metadata-based query, the Boolean matching technique is used to compare annotated metadata and user specified criteria. If any segment belonging to a news item matches the query, then all the other segments belonging to the news item are retrieved based on the sibling relationship, and these segments form a candidate set.

In an unstructured metadata-based query, the segments retrieved as a result of user specified criteria are clustered based on the similarity among the segments. Next, the clustered segments are augmented using a transitive search and the sibling relationships among the segments. The resulting clusters or candidate sets are used for compositions.

In composite metadata-based query, the common segments retrieved from the two individual queries (annotated metadata-based and unstructured metadata-based) are retained for composition.

The CGI scripts are written in the C language to execute the queries. The transitive retrieval technique and all the composition techniques are implemented as CGI scripts. The conceptual compositions formed from the candidate sets are displayed in the Web interface, from where the user initiates video playout. Video data are streamed separately through TCP/IP protocol and displayed using MTV, a MPEG-1 video playout client.

# Chapter 8

# Conclusions and Future Work

## 8.1 Conclusions

Evaluation of a video-based composition is a complex process as various features associated with a video composition need to be analyzed. While the existing metrics evaluate the retrieval performance of a DVPS, they are not useful in assessing the quality of a video composition. In addition, the existing automatic video composition techniques are based only on content, and do not consider the creation time and structure of a video piece. In this dissertation we have proposed a set of metrics for evaluation of quality of news video compositions. We have also proposed various automatic video composition and customization techniques that overcome the limitations of the existing methods. We used our proposed metrics to evaluate the quality of manually composed broadcast news, and obtain values that serve as references to judge the quality of an automatic newscast composition produced by our proposed composition techniques.

In this dissertation, we have introduced the concept of period-based and instance-based compositions. Period-based composition includes temporal ordering, thematic ordering, and thematic nearness ordering techniques. Temporal composition provides temporal ordering of a composition, but depends on similarity among the segments for thematic flow in the composition. Thematic composition maintains both correct temporal ordering and a smooth flow of information. However, compositions resulting from this technique can have large temporal jumps, either because the threads with considerable varying information are dropped, or because segments with significantly different creation times are considered similar. To overcome these shortcomings, we have proposed the thematic nearness composition technique. In this technique, the similarity between two segments is not only based on the information, but also on the difference in their creation time. We find that better overall composition quality is achieved as we move from the use of temporal to thematic nearness techniques.

In instance-based composition, we have used random ordering, clustering, and thematic ordering of the segments. In random ordering and clustering, we have assumed that the ordering of segments in a body is irrelevant if they belong to the same instance of time. Therefore, the segments in the body can be presented in any order or clustered based on their type and the clusters presented in any order. However, thematic ordering of body segments can be used in an instance-based composition to improve the smoothness in information flow in a composition.

In addition, we have proposed novel breadth-first & depth-second composition techniques for composition under playout time constraints. These techniques provide diversity in information and cover the maximum possible creation time period. Using these techniques on news video data, we find that though the information conveyed by the customized composition is reduced, as expected, the creation period covered is increased. Other composition features maintain reasonable quality as compared with broadcast news video composition.

We have also investigated other components of a news DVPS in order to implement a working digital news video production system (Section A1). We have proposed and defined metadata types, a concept ontology, and a concept/object model. We used these to develop an annotation engine for semi-automatic information extraction. We have also investigated information semantics to develop a hybrid technique for better recall and precision of the retrieval. We found a significant increase (48%) in recall when the proposed retrieval technique was used on our experimental data set.

## 8.2 Future Work

Our work can take a number of new directions. In particular, manipulation of information within segments is an area that needs to be explored. The manipulation of information within a segment can involve the introduction of a new object or the removal of an existing object from a segment. Manipulation of information in a segment will not only allow us to create a segment with added information, it will also permit customization of the information contained in the segment. The composition techniques proposed in this dissertation are based on temporal sequencing of video segments. Additionally, *scene composition* can also be investigated as a potential technique.

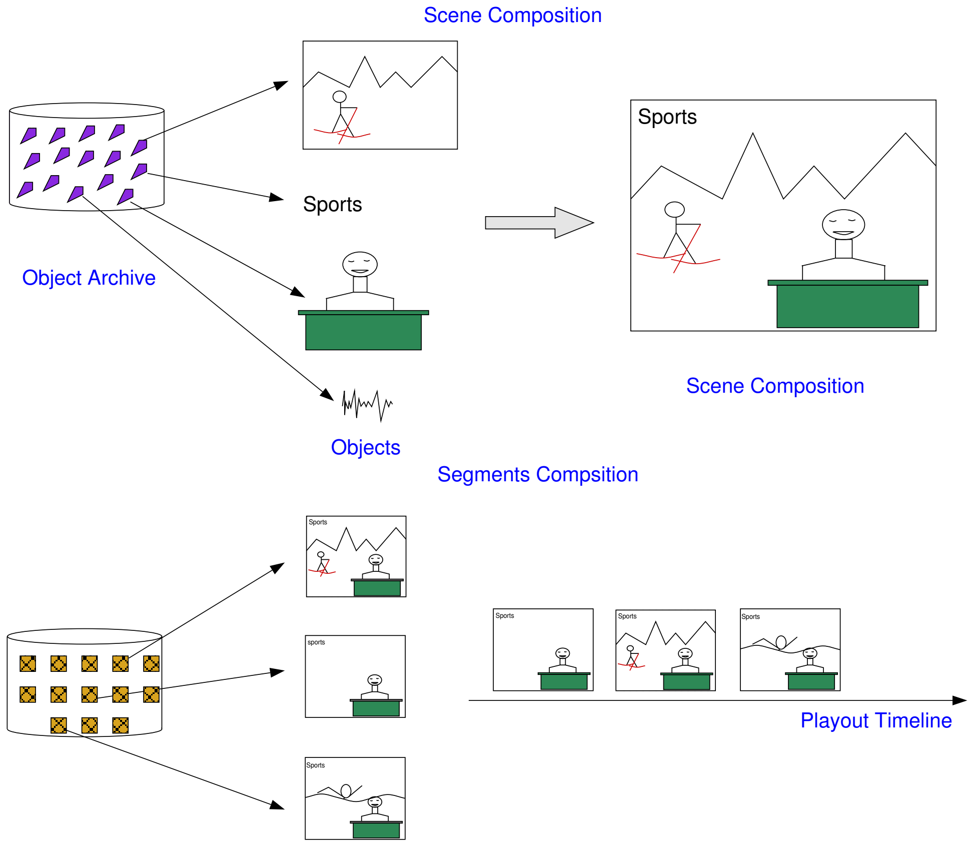


Figure 8.1: Automatic Scene Composition Concept

As shown in Fig. 8.1, given a set of objects (e.g., background, anchor, text, audio), we can automatically compose the objects to create a scene. Similarly, all the scenes required in a composition can be created. Currently, scenes are already pre-composed and stored in a video archive. The objectives of automatic scene composition is to achieve a dynamic and visually rich composition of a newscast. A visually rich composition can be achieved by a selection of interesting and informative video objects from any composition to create a new composition. It provides interactivity with objects within a scene that is otherwise not possible in simple playout of a video. Also, the same objects can be reused to create different scenes. We define the following two types of scene composition processes:

**Aggregate Scene Composition:** A composite of independent but related objects (Fig. 8.1).

**Partial Scene Composition:** Replacing objects in a composite. E.g., replacing a talking head with location scene.

Objects in a scene can be dropped or replaced only if visual objects are available and techniques to form a composite are available. MPEG-4 [59, 60], a digital video standard, can be used for the above objective. In this standard the concept of audio/visual objects (AVO) is present. Information about the objects and how these objects are to be rendered for final presentation can be incorporated in the MPEG-4 stream. In addition, AVOs can be natural or synthetic, (i.e., recorder with a camera/microphone or generated using a computer). Therefore, the MPEG-4 standard can be effectively used for aggregate and partial scene composition (Fig. 8.2) by incorporating information as to how the objects should be rendered.

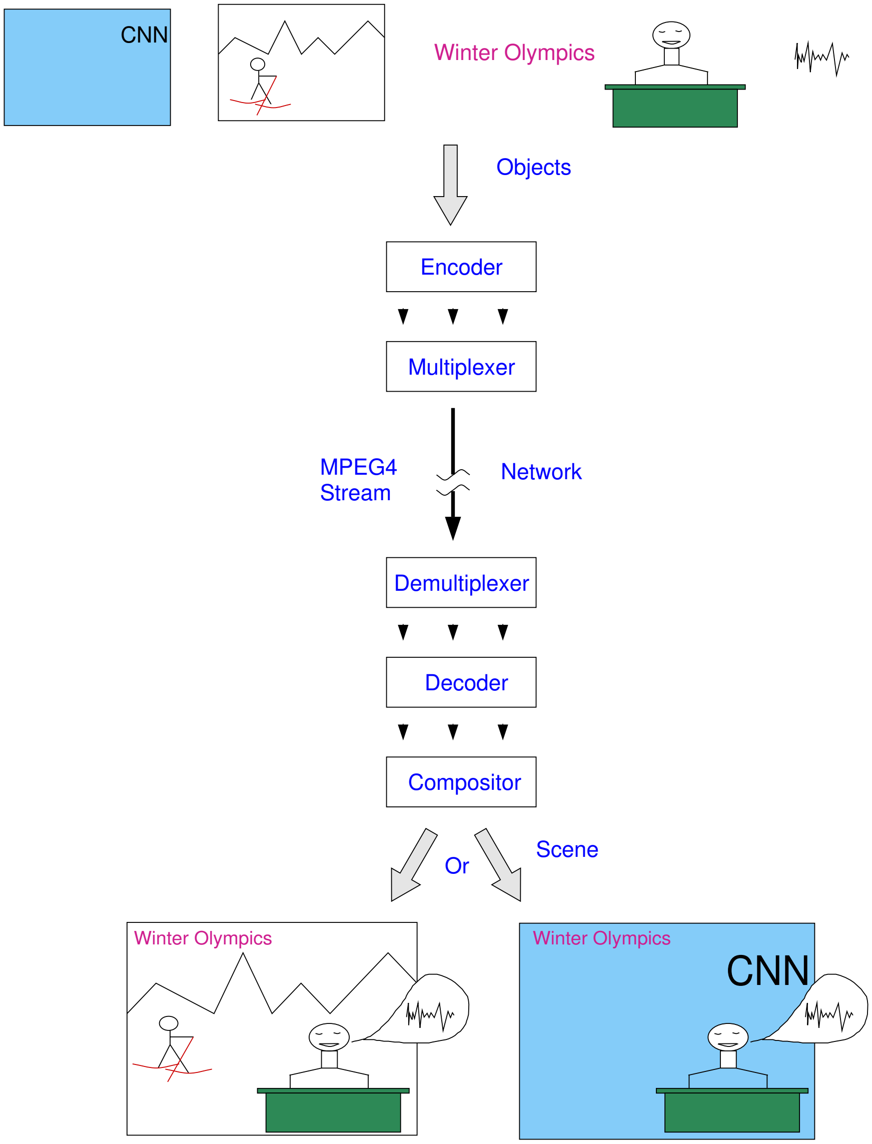


Figure 8.2: Automatic Scene Composition in MPEG4

The coding of objects separately also offers scope for content personalization at the object level in a video. In current composition techniques, we drop a complete video segment because of an undesired object within a segment. By encoding objects separately we can drop only the undesired objects and use rest of the objects in the segment composition. This is especially useful in resource management (e.g, network bandwidth). Currently the concept of streaming two different video compositions to two users with distinct requirements is used. By using dynamic object composition, multiple streaming can be avoided to a certain extent. It is possible to distribute a single stream but provide different composition to the users. All the objects desired by the users can be incorporated in a MPEG-4 stream and the desired objects can be incorporated at the client-end.

Therefore, multiple objects can be encoded and only partial objects can be decoded depending on a user's choice. Therefore, streams do not need to be encoded for individuals, and a single stream can be multicasted to different users. However, MPEG-4 compositor needs to be extended to incorporate user profile. Further, objects can be encoded at different resolutions depending on their importance.

Instead of storing metadata in a separate archive, MPEG-7 [61] defines a standard for description of multimedia content. This description can be attached to objects regardless of their format. Therefore, using MPEG-4, it is possible to attach descriptions (e.g., name of the person, date of an event, and location) with objects. This provides a convenient way of locating content in a video stream.

Using MPEG-4 and MPEG-7 encoding, the following objectives can be achieved in broadcast news composition:

1. Replacement of interesting visual content (e.g., location shots) with *dead* visual content (e.g., talking head).

2. Object personalization, depending on a user profile, irrelevant data can be dropped. Or, desired data are incorporated in a composition.

3. Enrichment of content by including "added value" information.

4. Object archive will be smaller due to reuse of content.

In summary, to provide further flexibility in video composition, techniques for composition and customization at scene level need to be investigated and developed. In addition, the metrics proposed in this work are based on the feature set specific to a news video domain. However, different subsets of the feature set can be used to evaluate other domain-specific compositions. For example, in the jokes domain the creation time of the content does not affect the information conveyed by a composition. Therefore, we need to identify subsets of the feature set for evaluation of other domain-specific compositions and identify, if any, additional domain-specific features required for evaluation.

# Bibliography

[1] G. Ahanger and T.D.C. Little, "Data Semantics for Improving Retreival Performance of Digital News Video Systems," to appear in 8th IFIP 2.6 Working Conference on Database Semantics, Rotorua, New Zealand, January 1999.

[2] G. Ahanger and T.D.C. Little, "Automatic Composition Techniques for Video Production," to appear in *IEEE Transactions on Knowledge and Data Engineering*, Vol. 10, No. 6, 1998.

[3] G. Ahanger and T.D.C. Little, "Automatic Digital Video Production Concepts," *Handbook on Internet and Multimedia Systems and Applications*, CRC Press, Boca Raton, December 1998.

[4] G. Ahanger and T.D.C. Little, "A Language to Support Automatic Composition of Newscasts," *Computer and Information Technology*, Vol. 6, No. 3, 1998, pp. 297-310.

[5] G. Ahanger and T.D.C. Little, "Easy Ed: An Integration of Technologies for Multimedia Education," *Proc. of WebNet '97*, Toronto,

Canada, October 1997.

[6] G. Ahanger and T.D.C. Little, “A System for Customized News Delivery from Video Archives,” *Proc. Intl. Conf. on Multimedia Computing and Systems*, Ottawa, Canada, June 1997, pp. 526-533.

[7] G. Ahanger and T.D.C. Little, “A Survey of Technologies for Parsing and Indexing Digital Video,” *Journal of Visual Communication and Image Representation*, Vol. 7, No. 1, March 1996, pp. 28-43.

[8] G. Ahanger, D. Benson, and T.D.C. Little, “Video Query Formulation,” *Proc. IS&T/SPIE Conf. on Storage and Retrieval for Image and Video Databases*, Vol. 2420, February 1995, pp. 280-291.

[9] A.V. Aho, R. Sethi, and J.D. Ullman, “Syntax-Directed Translation,” *Compilers: Principles, Techniques, and Tools*, Addison-Wesley Publishing Company, Reading, Massachusetts, March 1988, pp. 279-342.

[10] A. Akutsu and Y. Tonomura, “Video Tomography: An Efficient Method for Camerawork Extraction and Motion Analysis,” *Proc. ACM Multimedia ’94*, San Francisco, CA, 1994, pp. 349-356.

[11] J.F. Allen, “Maintaining Knowledge about Temporal Intervals,” *Comm. of the ACM,* November 1983, Vol. 26, No. 11, pp. 832-843.

[12] E. Ardizzone and M. La Casia, “Automatic Video Database Indexing and Retrieval,” *Multimedia Tools and Applications*, Vol. 4, No. 1, 1997, pp. 29-56.

[13] F. Arman, A. Hsu, and M-.Y. Chiu, "Image Processing on Compressed Data for Large Video Databases," *Proc. 1st ACM Intl. Conf. on Multimedia*, Anaheim CA, August 1993, pp. 267-272.

[14] F. Attneave, "Dimensions of Similarity," *Americal Journal of Psychology*, Vol. 63, 1950, pp. 516-556.

[15] A.D. Bimbo, M. Campanai, P. Nesi, "A Three-Dimensional Iconic Environment for Image Database Querying," *IEEE Trans. on Software Engineering*, Vol. 19, No. 10, 1993, pp. 997-1011.

[16] M. J. Black and Y. Yacoob "Tracking and recognizing facial expressions in image sequences, using local parameterized models of image motion," *Proc. Intl. Conf. on Computer Vision*, Cambridge, MA, 1995, pp. 374-381.

[17] K. Böhm and T.C. Rakow, "Metadata for Multimedia Documents," *SIGMOD Record*, Vol. 23, No. 4, December 1994, pp. 21-26.

[18] G. Bordogna, I. Gagliardi, D. Merelli, P. Mussio, M. Padula, and M. Protti, "Iconic Queries on Pictorial Data," *Proc. IEEE Workshop on Visual Languages*, October 1989, pp. 38-42.

[19] E. Branigan, *Narrative Comprehension and Film*, Rutledge, New York, 1992.

[20] 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. ACM Multimedia ’95*, San Francisco, CA, 1995, pp. 35-43.

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

[22] M. Carrer, L. Ligresti, G. Ahanger, and T.D.C. Little, “An Annotation Engine for Supporting Video Database Population,” Multimedia Tools and Applications Vol. 5, No. 3, November 1997, pp. 233-258.

[23] N.S. Chang, and K.S. Fu, “Picture Query Languages for Pictorial Data-Base Systems,” *Computer*, Vol. 14, No. 11, November 1981, pp. 23-33.

[24] S. -F. Chang, J.R. Smith, M. Beigi, and A. Benitez, “Visual Information Retrieval from Large Distributed Online Repositories,” *Comm. of the ACM*, Vol. 40, No. 12, 1997, pp. 63-72.

[25] F. Chen, M. Hearst, J. Kupiec, J. Pedersen, and L. Wilcox, “Metadata for Mixed-Media Access,” *SIGMOD Record*, Vol. 23, No. 4, December 1994, pp. 64-71.

[26] T. Clark, “WebTV Eyes Fall Launch,” *Interactive Week*, June 13, 1996.

[27] *Approximation Algorithms for NP-Hard Problems*, D. Hochbaum (editor), PWS Publishing, Boston, 1997, pp. 46-93.

[28] K. Compton and P. Bosco, "CNN Newsroom on the Internet: A Digital Video News Magazine and Library," *Proc. Intl. Conf. on Multimedia Computing and Systems*, Washington D.C., May 1995, pp. 296-301.

[29] T.H. Cormen, C.E. Leiserson, and R.L. Rivest, *Introduction to Algorithms*, MIT Press, 1990.

[30] G. Davenport, T.A. Smith, and N. Pincever "Cinematic Primitives for Multimedia", *IEEE Computer Graphics and Applications*, July 1991, pp. 67-74.

[31] G. Davenport and M. Murtaugh, "ConText: Towards the Evolving Documentary," *Proc. ACM Multimedia '95*, San Francisco, November 1995, pp. 381-389.

[32] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 1. No. 2, April 1979.

[33] M. Davis, "Media Streams: An Iconic Visual Language for Video Annotation," *Proc. IEEE Symposium on Visual Languages*, Bergen, Norway, 1993, pp. 196-202.

[34] R.O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, 1973.

[35] C. Faloutsos, "Access Methods for Text," *Computing Surveys*, Vol. 17, No.1, March 1985, pp. 49-74.

[36] W.I. Grosky, F. Fotouhi, and I.K. Sethi, "Using Metadata for the Intelligent Browsing of Structured Media Objects," *SIGMOD Record*, Vol. 23, No. 4, December 1994, pp. 49-56.

[37] J. Hafner, Harpreet Sawney, W. Equitz, M. Flickner, and W. Niblack, "Efficient Color Histogram Indexing for Quadratic Form Distance Functions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 1, No. 7, 1995, pp. 729-736.

[38] T. Hamano, "A Similarity Retrieval Method for Image Databases Using Simple Graphics," *IEEE Workshop on Languages for Automation, Symbiotic and Intelligent Robotics*, University of Maryland, August 29-31, 1988, pp. 149-154.

[39] A. Hampapur, R. Jain, and T. Weymouth, "Digital Video Segmentation," *Proc. ACM Multimedia '94*, 1994, pp. 357-364.

[40] J.A. Hartigan, *Clustering Algorithms*, John Wiley & Sons, 1975.

[41] K. Hirata and T. Kato, "Query By Visual Example," *Proc. Third Intl. conf. on Extending Database Technology*, Viennna, Austria, March 1992, pp. 56-71.

[42] E. Hyden and C. Sreenan, "Agora – A Personalized Digital Newsfeed," *Proc. 6th Intl. Workshop on Network and Operating Systems Support for Digital Audio and Video*, Zushi, Japan, Short Papers, April 1996.

[43] H.V. Jagdish, "A Retrieval Technique for Similar Shapes," *Proc. ACM SIGMOD*, May 1991, pp. 208-217.

[44] T. Joseph and A.F. Cardenas, "PICQUERY: A High Level Query Language for Pictorial Database Management," *IEEE Trans. on Software Engineering*, Vol. 14, No. 5, May 1988, pp. 630-638.

[45] J. Kamahara, S. Shimojo, A. Sugano, T. Kaneda, H. Miyahara, T. Yahata, and S. Nishio, "A News on Demand System with Automatic Program Composition and QOS Control Mechanism," *Intl. Journal of Information Technology*, Vol. 2, No. 1, 1996, pp. 1-22.

[46] W. Klaus and A. Sheth, "Metadata for Digital Media: Introduction to the Special Issue," *SIGMOD Record*, Vol. 23, No. 4, December 1994, pp. 19-20.

[47] W. Klippgen, T.D.C. Little, G. Ahanger, and D. Venkatesh, "The Use of Metadata for the Rendering of Personalized Video Delivery," in *Multimedia Data Management; Using Metadata to Integrate and Apply Digital Media*, Amit Sheth and Wolfgang Klas (eds.), McGraw Hill, March 1998, pp. 287-318.

[48] A. Kobsa and W. Pohl, "The User Modeling Shell System BGP-MS," *User Modeling and User-Adapted Interaction*, Vol. 4, No. 2, 1995, pp. 59-106.

[49] S.Y. Lee and H.M. Kao, "Video Indexing – An Object Based on Moving Object and Track," *Proc. IS&T/SPIE*, Vol. 1908, 1993, pp. 25-36.

[50] R. Lienhart, S. Pfeiffer, and W. Effelsberg, “Video Abstracting,” *Comm. of the ACM*, Vol. 40, No. 12, 1997, pp. 55-62.

[51] T.D.C. Little, G. Ahanger, R.J. Folz, J.F. Gibbon, A. Krishnamurthy, P. Lumba, M. Ramanathan, and D. Venkatesh, “Selection and Dissemination of Digital Video via the Virtual Video Browser,” *Journal of Multimedia Tools and Applications*, Vol. 1 No. 2, June 1995, pp. 149-172.

[52] T.D.C. Little and A. Ghafoor, “Interval-Based Temporal Models for Time-Dependent Multimedia Data, *IEEE Trans. on Knowledge and Data Engineering,* Vol. 5, No. 4, August 1993, pp. 551-563.

[53] S. Luke, L. Spector, D. Rager, and J. Hendler, “Ontology-Based Web Agents,” *Proc. 1st Intl. Conf. on Autonomous Agents*, Marina Del Rey, CA, 1997, pp. 59-66.

[54] P. Maes, “Agents that Reduce Work and Information Overload,” *Comm. of the ACM*, Vol. 37, No. 7, July 1994, pp. 31-40.

[55] T.W. Malone, K.R. Grant, F.A. Turbak, S.A. Brobst, and M.D. Cohen,“Intelligent information sharing systems,” *Comm. of the ACM*, Vol. 30, No. 5, May 1987, pp. 390-402.

[56] R.S. Marcus, “Computer and Human Understanding in Intelligent Retrieval Assistance,” *Proc. 54th American Society for Information Science Meeting*, Vol. 28, October 1991, pp. 49-59.

[57] N. Mathe and J. Chen,"A User-Centered Approach to Adaptive Hypertext based on an Information Relevance Model," *Proc. 4th Intl. Conf. on User Modeling (UM'94)*, Hyannis MA, August 1994, pp. 107-114.

[58] miniSQL: A Relational Database, *http://www.hughes.com.au/*.

[59] MPEG Requirements Group, "MPEG-4 Overview," *ISO/IEC/JTC1/SC29 /WG11/N2196*, March 1998, Tokyo.

[60] MPEG Requirements Group, "MPEG-4 Applications," *ISO/IEC/JTC1/SC29 /WG11/N2195*, March 1998, Tokyo.

[61] MPEG Requirements Group, "MPEG-7 Context and Objectives," *ISO/IEC/ JTC1/SC29/WG11/N2207*, March 1998, Tokyo.

[62] MPEG Playback Software, *www.mpegtv.com*.

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

[64] F. Nack and A. Parkes, "The Application of Video Semantics and Theme Representation in Automated Video Editing," *Multimedia Tools and Applications*, Vol. 4, No. 1, January 1997, pp. 57-83.

[65] A. Nagasaka, and Y. Tanaka, "Automatic Video Indexing and Full-Video Search for Object Appearances," *Visual Database Systems, II*, Eds. E. Knuth, and L.M. Wegner, Elsevier Science Publishers B.V., 1992 IFIP, pp. 113-127.

[66] A.D. Narasimhalu, M.S. Kankanhalli, and J. Wu, "Benchmarking Multimedia Databases," *Multimedia Tools and Applications*, Vol. 4, No. 3, May 1997, pp. 333-356.

[67] V. Ogle and Stonebreaker, "Chabot: Retrieval from a Relational Database of Images," *Computer*, Vol. 28, No. 2, 1995.

[68] E. Oomoto and K. Tanaka, "OVID: Design and Implementation of a Video-Object Database System," *IEEE Trans. on Knowledge and Data Engineering*, Vol. 5, No. 4, August 1993, pp. 629-643.

[69] J.A. Orenstein, and F.A. Manola, "PROBE Spatial Data Modeling and Query Processing in an Image Database Application," *IEEE Trans. on Software Engineering*, Vol. 14, No. 5, pp. 661-629, May 1988.

[70] J. Orwant, "Heterogeneous Learning in the Doppelgänger user Modeling System," *Journal of User Modeling and User-Adapted Interaction*, 1995.

[71] J. Ozer, "Industry Trends: Going Digital," *PC Magazine*, April 21, 1997.

[72] G. Ozsoyoglu, V. Hakkoymaz, and J. Kraft, "Automating the Assembly of Presentation for Multimedia Data," *IEEE Intl. Conf. on Data Engineering*, February 1996, pp. 593-601.

[73] A. Pentland, R.W. Picard, and S. Sclaroff, "Photobook: Content-Based Manipulation of Image Databases," *Proc. IS&T/SPIE Conf. on Stor-*

*age and Retrieval for Image and Video Databases II*, Vol. 2185, Febuary 1994.

[74] R. Picard and T. Minka, "Vision Texture for Annotation", *Multimedia Systems*, Vol. 3, No. 3, 1995, pp. 3-14.

[75] M. Rabiger, *Directing the Documentary*, 3rd Ed., Focal Press, Boston, 1998.

[76] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," *Proc. CSCW '94*, Chapel Hill, NC, October 1994, pp. 175-186.

[77] L.A. Rowe, J.S. Boreczky, and C.A. Eads, "Indexes for User Access to Large Video Databases," *SPIE*, Vol. 2185, Febuary 1994, pp. 150-161.

[78] W. Sack and M. Davis, "IDIC: Assembling Video Sequence from Story Plans and Content Annotation," *Proc. Intl. Conf. on Multimedia Computing and Systems*, Boston, Massachusetts, May 1994, pp.30-36.

[79] G. Salton and M.J. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill Book Company, New York, 1983.

[80] S. Santini and R. Jain, "Similarity is a Geometer," *Multimedia Tools and Applications*, Vol. 5, No. 3, 1997, pp. 277-306.

[81] S. Sclaroff and J. Isidoro, "Active Blobs," *Proc. Intl. Conf. on Computer Vision*, Mumbai, India, 1998.

[82] B. Shahrary and D. Gibbon, "Automatic Generation of Pictorial Transcripts of Video Programs," *Proc. Intl. Conf. on Multimedia Computing and Networking, SPIE*, San Jose, February 1995, pp. 512-518.

[83] U. Upendra Shardanand and P. Maes,"Social Information Filtering: Algorithms for Automating Word of Mouth," *Proc. ACM CHI '95*, 1995, pp. 210-7.

[84] B.D. Sheth, "A Learning Approach to Personalized Information Filtering," *MS Thesis*, MIT, Cambridge MA, February 1994.

[85] G. Silva and C.A. Montgomery, "Knowledge Representation for Automated Understanding of Natural Language Discourse," *Computer and Humanities*, Vol. 11, No. 4, July 1977, pp. 223-234.

[86] L. Simone, "Video Editing," *PC Magazine*, October 7, 1997.

[87] M.A. Smith and T. Kanade, "Video Skimming and Characterization using a combination of Image and Language Understanding," *Proc. IEEE Intl. Workshop on Content-Based Access of Image and Video Databases*, Bombay, India, January 1998.

[88] T.G. Aguierre Smith, and G. Davenport, "The Stratification System: A Design Environment for Random Access Video," *Proc. 3rd Intl. Workshop on Network and Operating System Support for Digital Audio and Video*, La Jolla, CA, November 1992.

[89] K. Tsuda, M. Hirakawa, M. Tanaka, and T. Ichikawa, "IconicBrowser: An Iconic Retrieval System for Object-Oriented Databases," *Proc. IEEE Workshop on Visual Languages*, Italy, October 1989, pp. 130-137.

[90] J.S. Wachman, "A Video Browser that Learns by Example," Master Thesis, Technical Report #383, MIT Media Laboratory, Cambridge, Massachusetts, 1997.

[91] H. Wactlar, T. Kanade, M.A. Smith, and S.M. Stevens, "Intelligent Access to Digital Video: The Informedia Project," *Computer*, Vol. 29, No. 5, 1996, pp. 46-52.

[92] R. Weiss, A. Duda,and D.K. Gifford, "Composition and Search with a Video Algebra," *IEEE Multimedia*, Spring 1995, pp. 12-25.

[93] T.W. Yan, M. Jacobsen, H. Garcia-Molina, and U. Dayal, "From User Access Patterns to Dynamic Hypertext Linking," *Computer Networks and ISDN Systems,* Vol. 28, No. 7, pp. 1007-14.

[94] H.J. Zhang, A. Kankanhalli, and S.W. Smoliar, "Automatic Partitioning of Full-Motion Video," *ACM/Springer Multimedia Systems*, Vol. 1, No. 1, 1993, pp. 10-28.

# Appendix A1

# Summary of Requirements and Techniques for a Video Production System

Here we summarize the requirements and the implementation techniques for a video production system. This section includes techniques used for query matching with annotated and unstructured metadata; video data retrieval techniques including transitive search and union based search; and techniques used for composition of a video piece.

## A1.1 Requirements:

1. Digitized video segments must be available. Information contained in a video segment should be complete and self contained. Each segment

should have minimal dependencies on neighboring segments (Section 4.1.2).

2. Closed-caption data associated with the video segments must be available. Closed-caption data associated with each video segment are treated as metadata associated with the segment (Section 7.2.2).

3. Information about events, places, and persons in the visuals (Section 7.2) must be available, potentially via human annotation.

4. Information about the sibling relationships (Sections 7.2 and 6.3) must be available.

5. Annotated information about the creation time and the playout duration of each segment (Sections 7.2 and 6.3) must be available.

6. Database to manage metadata about segments must be available. In the implementation of Canvass we used a relational database called miniSQL [58].

7. Two set of indices generated from the closed-caption/unstructured data for initial retrieval and clustering/transitive search (Sections 7.2.2 and 6.4) must be available.

8. Tools to index the unstructured data must be available. In the implementation of Canvass we used SMART [21] to create the indices.

Next, we discuss the techniques for the implementation of a video production system.

## A1.2 Implementation

In this section we discuss the video data retrieval and composition techniques.

**Retrieval Techniques:** Video segments can be retrieved by matching either or both annotated metadata and unstructured metadata (Section 7.3).

The resultant segments retrieved as a result of any of the above matching techniques will be part of different storylines. Therefore, further processing is required to separate the segments into different storylines; therefore, the segments need to be clustered (Section 6.4). Each cluster represents a single storyline. The recall of the system can also be increased as follows (Section 7.3):

1. Use a minimal spanning tree technique [29, 34] for clustering segments into separate storylines.

2. After the clusters are formed, these clusters can be augmented as follows:.

   (a) First, use a transitive search technique for each segment in a cluster. The additional segments retrieved in the process are retained as elements of the cluster.

   (b) Second, use a sibling search technique for each segment in a cluster. The additional segments retrieved in the process are again retained as elements of the cluster.

**Composition:** The following techniques can be used to compose the segments in each cluster into a narrative (Section 4.2):

- For instance-based composition we must maintain the structural and thematic continuity.

    1. Select a segment of type Introduction.
    2. Order the segments in the body in a random order, cluster according to the segments types and order, or using cosine similarity for ordering.
    3. Select a segment of type Enclose, if available.

- For period-based composition we must maintain the structural, temporal, and thematic continuity.

    1. Select a segment of type Introduction with earliest creation time and date.
    2. Order the segments in the body in a chronology.
    3. If better thematic continuity is desired use cosine similarity.
    4. Select a segment of type Enclose with the latest creation time and date.

**Time-Limited Composition:** Constraints are imposed on the playout duration of a composition and following two situations can occur (Section 4.2.3):

- If more playout time is available than the total playout time of a set of compositions then insert value-added content (e.g., advertisements).
- If less playout time is available (Section 4.2.3) than the total playout time of a set of compositions then drop data as follows:
    - Instance-Based: Fit in as many views and utilize all the available playout time as follows:
        1. Distribute the available playout time among the composition by proportionally dividing the playout time according to the playout time of each composition.
        2. Cluster the segments in the body of a composition according to their type and select a segment from each cluster and keep iterating until no more time can be utilized for the composition.
        3. In the end if some time is left for the composition that could not be adjusted, then accumulate the time.
        4. Now the objective is to utilize as much of the accumulated time. Therefore, use bin packing technique for selecting segments.
    - Period-based: Optimize the span covered in a composition and utilize all the available playout time as follows:
        1. Distribute the available playout time among the composition by proportionally dividing the playout time according to the

playout time of each composition.

2. Divide the creation time line of each composition into sub-periods. Select a segment from each sub-period and keep iterating until no more time can be utilized for the composition.
3. In the end if some time is left for the composition that could not be adjusted, then accumulate the time.
4. Now the objective is to utilized as much of the accumulated time. Therefore, use bin packing technique for selecting segments.

# Appendix B1

# Glossary

**Candidate Set:** A set of segments selected by a query on a data universe.

**Canonical Model:** Formal encoding of an application-specific cognitive/semantic user profile.

**Clip:** Same as a segment.

**Composition:** The process of sequencing video segments to create a narrative.

**Composition Set:** A set of segments that are part of a composition.

**Concept Vector:** A vector that consists of concepts associated with video data.

**Content Progression:** Rate of change in information within a composition.

**Cover Footage:** Video segments that encompass all of the aspects of a story center. For example, shots from the scene, comments of by-standers, and formal interview about the story.

**CM:** Continuous Media (e.g., video).

**Creation Time:** Date and time when video is recorded.

**Customization:** Tailoring a narrative according to user or system constraints.

**Data Model:** Representation of extracted information from video data and the relationships.

**DTD:** Document type definition; contains context rules of an SGML document.

**DVPS:** Digital video production system.

**Event:** Anything that happens; an occurrence of some importance in a certain place during a particular interval of time.

**FCC:** Federal Communications Commission.

**Focus:** The main concept in a narrative.

**Historical Time:** Creation time.

**Narrative:** A narrative is a series of events collected as a cause and effects chain.

**Ontology:** Description of the concepts and relationships that can exist in video data.

**Period-Span Coverage:** The span encompassing the life of an event.

**Precision:** Measurement of the ability of the system to present relevant data.

**QBE:** Query by example.

**Information:** Concepts associated with video segments.

**IQ:** Iconic query.

**Recall:** The ability of the system to retrieve all relevant data.

**SGML:** Standard generalized markup language; a markup language used to define the structure of and manage documents in electronic form.

**Segment:** A shot or contiguous collection of shots forming whole unit of information.

**Semi-Structured Metadata:** Information contained in transcripts associated with video segments.

**Shot:** One or more frames recorded contiguously and representing a continuous action in time and space.

**Similarity Distance:** Separation in concepts between any two segments.

**SQL:** Structured query language.

**Story Center:** A focus.

**Story line:** A narrative

**Structural Continuity:** Metric the position of a segment type in a composition.

**Sub-event:** Cause and effects in an event over an interval of time.

**Tag:** Ending segments in a composition (enclose)

**Time Constraint Composition:** Composition achieved under playout time restrictions.

**Temporal Continuity:** Metric the sequencing of segments in time.

**Thematic Continuity:** Metric the smooth flow of information between consecutive segments.

**Thematic Jump:** Similarity distance.

**Theme:** A focus.

**Thread:** Temporally-ordered segments that present information about an aspect of an event.

**Transitive Search:** The process of retrieving segments that are similar to the segments returned as result of a user query.

**User Profile:** Information about constraints of a user, (e.g., content, playout time, cost, etc.).

**Video Archive:** Video data storage system.

**Video:** A sequence of synchronised pictures and audio data.

**Visuals:** Pictures.

**Wild Scene:** Footage from the actual location of the event.

## Biography

Gulrukh Ahanger is a PhD candidate in the Department of Electrical and Computer Engineering at Boston University. She is currently a research assistant in the Multimedia Communications Laboratory. Her research interests include visual query and composition systems for video.

Ms. Ahanger received her BE degree in Electronics and Communications Engineering from the Regional Engineering College, Srinagar, India in 1988 and the MS degree in Systems Engineering from Boston University in 1993. She consulted as a software engineer for Siemens Medical Electronics, Danvers, Massachusetts from May 1992 to January 1993. She also interned at Siemens Corporate Research Inc. during the summer of 1993 where she developed a prototype video query system. She worked on building a prototype for Remote Patient Telecare System from June 1994 to May 1995. Currently she is involved in a project for fast access to multimedia information which includes distance learning and customized-news-service applications.