Automatic Composition Techniques for Video Production[1]

G. Ahanger and T.D.C. Little

Multimedia Communications Laboratory
Department of Electrical and Computer Engineering
Boston University, 8 Saint Mary's Street
Boston, Massachusetts 02215, USA
(617) 353-9877, (617) 353-6440 fax
{*gulrukh,tdcl*}*@bu.edu*

MCL Technical Report No. 12-30-1998

Abstract–Video production involves the process of capturing, editing, and composing video segments for delivery to a consumer. A composition must yield a coherent presentation of an event or narrative. This process can be automated if appropriate domain-specific metadata are associated with video segments and composition techniques are established. Automation leads to the support of dynamic composition and customization for applications such as news on demand.

In this paper we present techniques to achieve dynamic, real-time, and cohesive video composition and customization. We also identify metrics for evaluation our techniques with respect to existing manually-produced video-based news. The results of such an evaluation show that the quality of automatic composition is comparable to, and in some cases, better than, broadcast news video composition. The results also validate the assertions on which the automatic composition techniques are based.

Keywords: Video production, automatic composition, domain-specific structure, customization, temporal constraints, evaluation metrics, thematic continuity, temporal continuity, structural continuity.

[1] *IEEE Trans. on Knowledge and Data Engineering*, Vol. 10, No. 6, December 1998, pp. 967-987. This work is supported in part by the National Science Foundation under Grant No. IRI-9502702.

1 Introduction

The development of digital video-based communication systems is heavily influenced by the capabilities of new digital technology. High-density storage, low cost compression hardware, standard protocols, and ubiquitous networking have enabled a class of video-based applications that were not previously viable. These applications include desktop video conferencing using the Internet (e.g., distance learning), video asset management and scheduling (e.g., cable TV ad insertion), and video database applications (e.g., movie preview kiosks).

Common to these applications is the need to manage access to the inherently linear and time-dependent media of audio and video. This access is interesting both within (e.g., a scene of a movie) and across multiple instances of the medium (many movies). This type of access is analogous to keyword-based searching on text documents and document sets. Once this type access is achieved, additional manipulation of the audio/video media are possible to support customization of content and dynamic assembly.

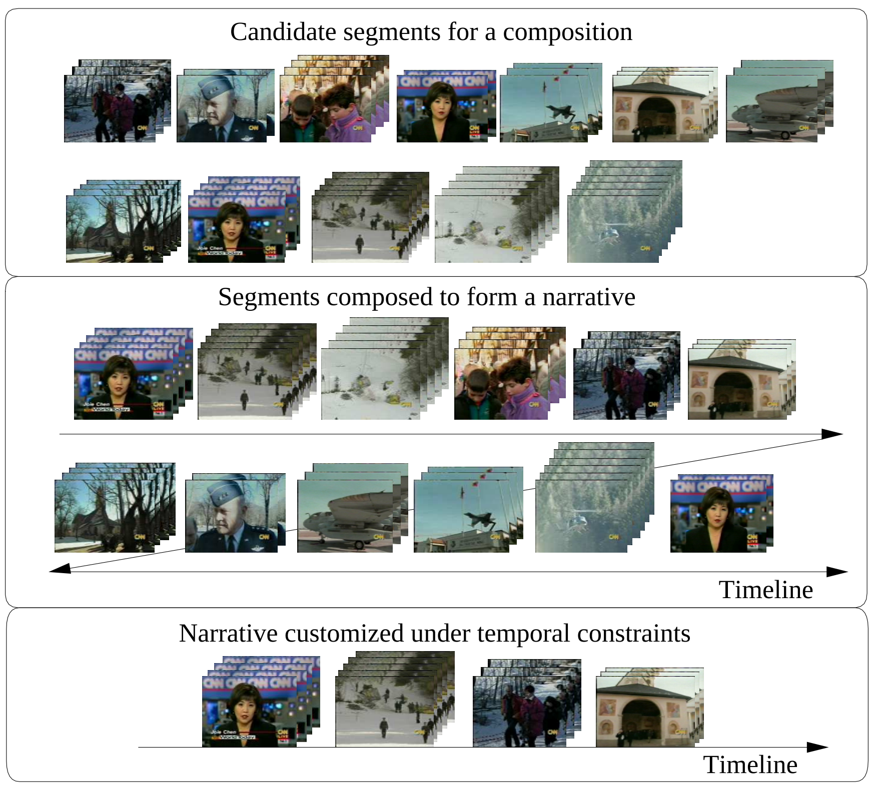

Figure 1: Example of Composition and Customization of Video Segments

In the digital domain, video data can be manipulated through automatic on-line selection, editing, assembly, and dissemination [1]. Video that have been used to create a movie or a news story in a single rendering can then be used in multiple contexts without involving an extensive re-production process. Simple access to multiple video repositories facilitates dynamically composing video segments to mechanically produce a narrative. Such an automation must yield a cohesive presentation that conveys a story.

Other possibilities include supporting replacement of obsolete video segments in a previously rendered video story. This is particularly applicable to video-based newscasts. Automatic establishment of relationships among the segments can also reduce authoring time. Therefore, dynamic video-based production can provide much more flexibility, ease of authoring, re-use, and customization for video applications. Fig. 1 illustrates an example of video composition beginning with a set of available video segments called a candidate set and proceeding to a video composition that conforms to temporal constraints. The process can be repeated to render multiple compositions, for example, as required to produce the multiple news items of a newscast. Typically the candidate set will be selected by a query on a data universe with parameters specific to the application that define the composition.

In this paper, we focus on techniques to compose a piece of video along a storyline with smooth flow of information and to customize the composition under temporal playout constraints. We apply these techniques to compose a digital video newscast.

Work related to automatic video composition includes references [6, 10, 12, 14]. Davenport et al. [6] maintain temporal continuity between segments by scoring the metadata associated with all available scenes. Weiss et al. [14] propose composition based on video algebra where the video model consists of hierarchical composition based on content or descriptive information associated with the segments. The segments are composed using algebraic operations like union, concatenation, and intersection. Nack and Parkes [10] establish continuity by rules based on the content. Ozsoyoglu et al. [12] compose multimedia presentations under temporal constraints by exclusion and inclusion of segments by also establishing content-based rules.

One main difference in our work is that we use knowledge of domain-specific structure and creation time in addition to content information. For example, news composition is highly dependent on creation time and the component segment types. We use these dependencies to yield compositions with correct domain-specific structure, correct time series of concepts, and acceptable thematic continuity.

In order to evaluate any composition technique, there is a need for metrics to quantify the quality of a composition. We seek to evaluate how well a dynamically-assembled composition performs with respect to a manually edited one. The metrics used for measuring *recall* and *precision* [13] remain valid for data retrieval; however, these metrics are oriented towards boolean evaluation (i.e., a retrieved object either matches a query or it does not).[2]

Ranked evaluation metrics can also be used to measure retrieval performance. A retrieved object may not exactly match the query but can have a degree of similarity. A rank-based metric can be applied to evaluate multimedia data retrieval. For example, if an image is retrieved, the degree of similarity between the query and the retrieved image can be measured

[2]Recall measures the ability of the system to retrieve all relevant data. Precision measures the ability of the system to present relevant data.

Table 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	kth set of composition segments (multiple compositions)

and ranked accordingly. Narasimhalu et al. [11] have proposed such metrics for retrieval of multimedia objects. These metrics characterize *rank, order, spread,* and *displacement* of retrieved objects. However, in our context of composition, we do not find application of these metrics.

In this paper we propose metrics that provide a comprehensive evaluation of the quality of an automatic video composition. The metrics are based on a feature set (e.g., time, theme, and structure) that completely represents a composition. We present automatic video composition techniques that incorporate the features as well as a new technique for composition of video under temporal constraints. Finally, the metrics are used to compare the quality of an automatic composition with manually-produced broadcast news and to verify the assumption on which the composition techniques are based.

The remainder of the paper is organized as follows. In Section 2 we present metrics for the evaluation of composed video. In Section 3 we describe rules and techniques used to compose and customize video-based newscasts. In Section 4 we present the evaluation of the proposed techniques. Section 5 concludes the paper.

2 Metrics for the Evaluation of Video Composition

To define suitable metrics we first understand the character of video as a medium for story telling, the nature of video composition, and the goals for comparing the quality of composition. We then define metrics and perform an evaluation of data collected from a specific video application domain. (The symbols used throughout the paper are summarized in Tables 1 – 4.) Later, in Section 4 these evaluation results are used as reference values to understand the performance of our composition algorithms.

2.1 Characteristics of the Video Medium

In a multimedia composition, the manner in which objects (e.g., text, graphics, video clips) are orchestrated is limited by an author's creativity and the constraints imposed by the capabilities of the multimedia rendering device [8]. However, video-based compositions are almost exclusively[3] comprised of sequentially-ordered video segments such that:

$$\forall i \; : \; 1 \; \leq \; i \; < \; N_c \; : \; s_i \text{ meets } s_{i+1},$$

where N_c is the total number of segments in an ordered set of segments of the composition $\{s_1, s_2, ..., s_{N_c}\}$. This ordering is based on the composer's intent and does not necessarily correspond to a chronology.

To convey a story using the video medium requires such a succession of video segments corresponding to concepts, or threads, of its narrative. A narrative has been defined in this context as a series of events collected as a chain [4]. 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 to concepts or threads that include story center and multiple related concepts. We use the terms "story center," "main concept," and "focus" interchangeably.

To create a video-based story one needs to have a description of the intended narrative and the parts of the narrative achieved by each video segment. Customization of a narrative or the corresponding video-based story requires appropriate filtering of the set of available video segments. This activity can be performed in ignorance of when the video segments are produced. For example, Fig. 2 illustrates this process in which structures in multiple time references are shuffled and mapped to a final story and playout timeline. To quantify the character of these shuffled video segments we identify some fundamental attributes, or the feature set, of video narratives.

The first 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 effects 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, in a news item multiple views of an event are presented (e.g., field shots and interviews). Therefore, there are different sets of segments

[3]Exceptions include navigation-based environments with branch points such as Lippman's Movie-Maps [7] and video games. However, in this paper we focus on linear video segment ordering.

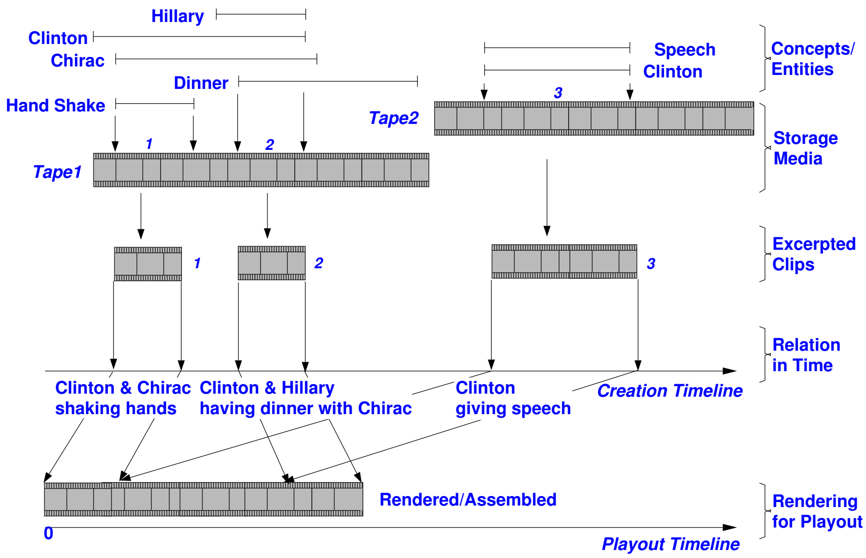

Figure 2: Composition Achieved by Shuffling of Segments in Time

Table 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

that present diverse information but by different vehicles. We related these sets possessing temporally-ordered segments to *threads*, where each thread contains information from different perspective about an event. Each a thread induces a shift in the theme of the story and continuity in theme needs to be maintained throughout a composition.

The lifespan of an event can vary from a day to years. A composition can encompass this entire period or a subset. We describe and quantify this coverage as *period span coverage*. We also consider continuity of the assembled components of the composition. For example, news video has structure consisting of an introduction, a body, and an end. A composed 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.

Therefore, the feature set for characterizing video compositions consists of information, thematic continuity, temporal continuity, structural continuity, period span coverage, and content progression. We define techniques to quantify these attributes next.

2.2 The Metrics

For each of the attributes in the feature set we propose a metric. The evaluation of the metrics assumes the existence of a *candidate set* S_a of segments for a composition. That is, the candidate set of segments satisfy a particular selection criterion from the universe of available video segments, S, and are relevant based on that 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 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 upon inclusion in S [5].

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

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

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

2.2.1 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 [13]. 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 observed in our data set 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, $In < 1$. An example of the application of this metric is provided in Appendix A.

2.2.2 Temporal Continuity

Temporal continuity, e_{tc}, is quantified as follows. Let N_c represent the number of segments placed on the creation timeline. The distances between segments are measured in time. We consider large forward jumps in time (if such data exist) to be less damaging to temporal continuity than reverse jumps. Therefore, we weigh forward jumps by $0 \leq \beta \leq 1$. Good temporal continuity means that all cause-effects in a story follow an increasing time series.

Temporal Continuity:

Table 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
β	Forward jump weight for temporal continuity
δ	Forward jump tolerance
λ	Dissimilarity threshold
τ	Similarity threshold
ρ	Fast change threshold
ϱ	Slow change threshold
D_t	Target temporal span
D_a	Achieved temporal span

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

Here, δ 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 $\frac{1}{N_c-1}\sum_{i=1}^{N_c-1} e_{tc}^i$. An example of temporal continuity evaluation is provided in Appendix A.

2.2.3 Thematic Continuity

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

Thematic Continuity:

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

The mean thematic continuity of a composition is $\frac{1}{N_c-1}\sum_{i=1}^{N_c-1} e_{thc}^i$. An example of thematic continuity evaluation is provided in Appendix A.

Table 4: 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
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

2.2.4 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.

Assuming that there are variations in the information contained in consecutive segments, and if the duration of playout of the consecutive segments is smaller than a fast-change threshold ρ, then we consider content progression as being fast. If the playout duration of a segment is greater than a slow-change threshold, ϱ, 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:

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

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 $\frac{1}{N_c}\sum_{i=1}^{N_c} e_{cp}^i$. An example of content progression evaluation is provided in Appendix A.

2.2.5 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 $\frac{D_a}{D_t}$. Appendix A contains an example of the application of this metric.

2.2.6 Structural Continuity

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

Structural Continuity for News Items:

$\{s_h\} = S_c \Rightarrow e_{sc} = 1$	Only a headline can be present in a composition set S_c.
$\{s_h, s_{in}\} = S_c \Rightarrow e_{sc} = 1$	Only a headline and an introduction can be present in a composition set S_c.
$\{s_{in}\} = S_c \Rightarrow e_{sc} = 1$	Only an introduction can be present in a composition set S_c.
$\{s_h, s_{in}, S_b\} = S_c \Rightarrow e_{sc} = 1$	Only a headline, an introduction, and segments belonging to the body can be present in a composition set S_c.
$\{s_h, S_b\} = S_c \Rightarrow e_{sc} = 1$	Only a headline and segments belonging to the body can be present.
$\{s_h, S_b, s_e\} = S_c \Rightarrow e_{sc} = 1$	Only an introduction, segments belonging to the body, and an enclose can be present.
$\{s_h, s_{in}, S_b, s_e\} = S_c \Rightarrow e_{sc} = 1$	All the segment types are present.
$\{\text{All other combinations}\} \Rightarrow e_{sc} = 0$	

In the above, $s_h \in S_h$, $s_{in} \in S_{in}$, $s_e \in S_e$, and S_b is a set with segments belonging to a body.

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. Later, these reference values are used to understand the performance of our automatic composition techniques.

2.3 Analysis of Broadcast News Video

Broadcast news video production presents us with well-defined domain-specific structures on which to apply our techniques. It is also readily available in ample quantities. In the following we describe our data collection and analysis including evaluation based on the aforementioned metrics for which we collected data.

2.3.1 News Video Data Collection

Broadcast news video data were acquired from CNN, NBC, and ABC over a period of 40 days from 20th January 1998 to 28th 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 [3] (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.

2.3.2 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 5.

The measurements show a thematic continuity that varies between 0.52 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.

2.3.3 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 temporal 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.

The minimum inter-transposition time found in the result data set (Table 6 and Fig. 3)

Table 5: Thematic Continuity and Content Progression Measurements

News Item	**No. of Segs**	e_{thc}	e_{cp}
1	9	0.80	0.91
2	11	0.93	0.93
3	8	0.89	0.97
4	7	0.89	1.0
5	5	0.68	0.87
6	2	1.0	0.88
7	10	0.94	0.93
8	8	1.0	1.0
9	6	0.99	1.0
10	7	0.73	0.96
11	5	0.98	0.92
12	7	0.98	1.0
13	5	1.0	0.94
14	5	0.92	0.92
15	3	0.85	0.87
16	6	0.94	1.0
17	3	0.52	1.0
18	10	0.98	0.93
19	6	0.99	0.95
20	4	1.0	0.90
21	4	0.98	0.86
22	7	1.0	0.96
23	5	1.0	1.0
24	6	0.95	0.88
25	2	1.0	0.81
26	4	1.0	1.0
27	8	0.97	1.0
28	8	0.97	0.92
29	8	0.90	0.89
30	8	0.97	0.98
31	6	1.0	0.92
32	6	1.0	0.95
33	4	0.86	0.93
34	5	0.95	0.95
35	6	1.0	1.0
36	4	1.0	0.93
37	6	1.0	1.0
38	3	0.95	1.0
39	10	1.0	0.95
40	7	1.0	0.94
41	3	1.0	1.0
42	5	1.0	0.95
43	5	0.98	1.0
44	7	1.0	1.0
45	11	0.98	1.0
46	9	1.0	1.0
47	11	1.0	0.95
48	4	0.94	1.0
49	12	1.0	0.93
50	7	0.98	0.98

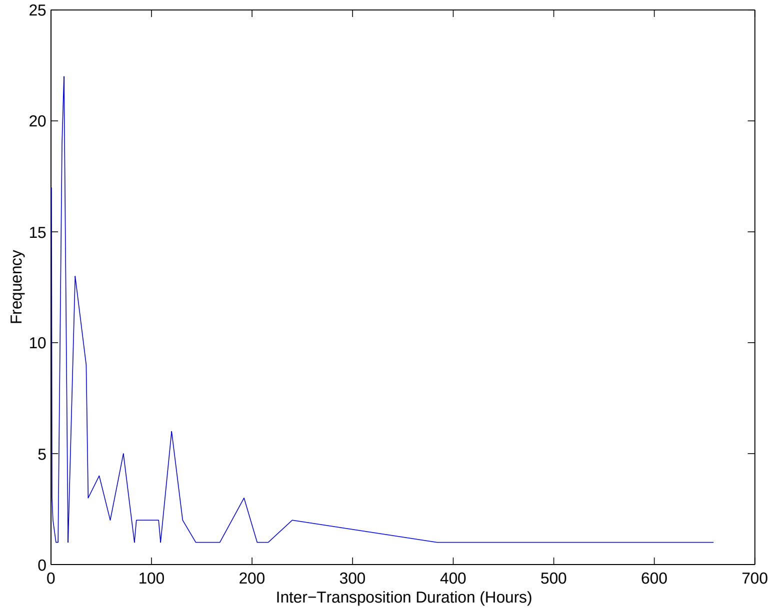

Figure 3: Inter-Transposition Durations for News Items from CNN1 and CNN2

is less than one hour (i.e., the same segment is repeated in a single broadcast). On average, the interval between a 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 6 shows additional data characterizing the other news sources.

Table 6: Segment Inter-transposition Repetition History

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

Fig. 4 illustrates the types (described in Section 3) 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 a subject; however, when the same segment

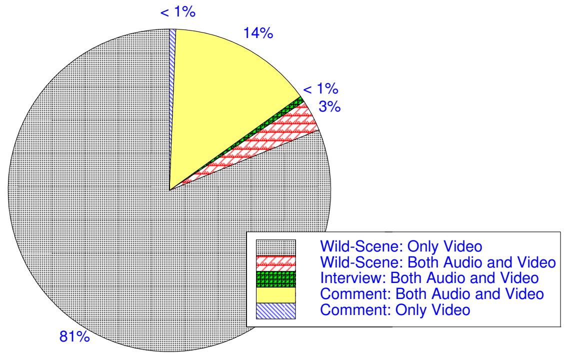

Figure 4: Segment Types Repeated by CNN

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 may 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 (CNN). In Table 7 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 7: Temporal Continuity Measurements

Parameter	**Value**
Presentation Duration	912 hrs
Total No. Segs	387
No. of Segs Repeated (Table 6)	17
Average Inter-Transposition Time (Table 6)	59 hrs
e_{tc} for Flashback	0.93
Tolerated Forward Jump Value δ	24 hrs
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 $1/386(15.8 + 16.6 + 352) = 0.99$.

2.3.4 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 fits 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 day 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 Composition and Customization Techniques

Recall our goal to achieve automated assembly of video segments to yield a cohesive video narrative. Up to this point we have characterized video compositions and established methods to quantify their attributes. We now introduce our proposed techniques to mechanize the video composition process. Because we require a domain-specific structure we begin with a description of the structure of video-based news. After establishing the basic segment types for this news domain we describe our composition techniques and approaches to composition under time constraints. The taxonomy of Fig. 5 illustrates the relationships among the proposed techniques described in this section. Table 8 summarizes the additional symbols used in this section.

3.1 Segments Types and Structure for the News Domain

We adopted the work of Musburger [9] as a reference structure for composition of a news item. Under this model (Table 9), 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.

Table 8: Symbols Used to Define Composition

Symbol	Description
N_{ni}	Number of news items in a newscast
$I(s)$	Interest associated with a segment
IS	Information similarity
CS	Content similarity
d_u	Target playout duration of a newscast
d_c	Playout duration of a newscast
d_{S_c}	Playout duration of a composition set
$d^r_{S_c}$	Duration yet to be accommodated during composition of a news item
TP_i	Sub-period on timeline
t	Time

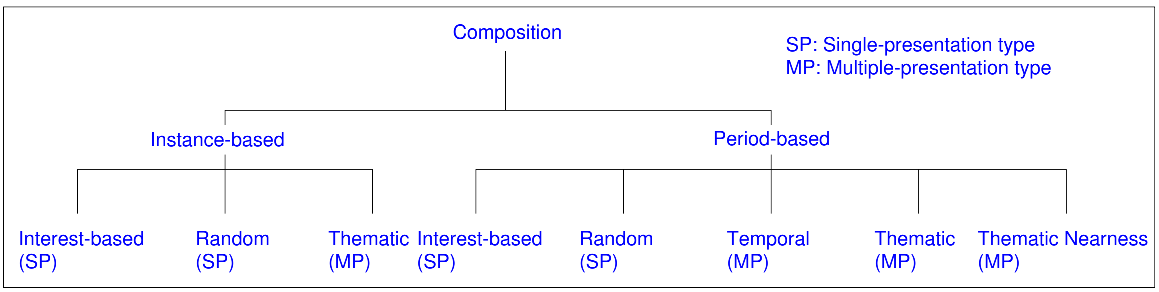

Figure 5: Taxonomy of Proposed Composition Techniques

As before, 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.

Table 9: Structure of a News Item

Headline		
Introduction		
Body (current)	Comment	
	Wild Scene	
	Interview	Question & Answer (QA)
	Speech	
	Enactment	
Enclose		

3.1.1 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). The single-presentation-type of segments are selected based on their importance value (instance-based) and creation time (period-based). We use different rules for each.

Creation Instance: 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. 6 illustrates this type of composition.

Creation Period: If a period is indicated then the rules specified in Table 10 are followed to select a single-presentation-type segment.

3.1.2 Multiple-Presentation Type

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

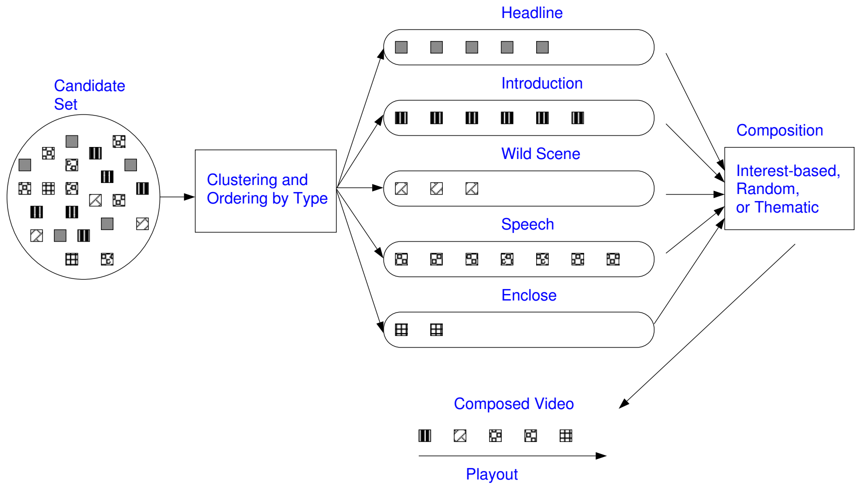

Figure 6: An Example of Instance-Based Composition

Table 10: 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 set with 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 Introduction 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 has 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 particular date then we use the segment s_k with the highest interest.

Creation Instance: Because there can be more than one segment mapping to the same instance on the the creation timeline, segments belonging to a body can be grouped (clustered) depending on their type (i.e., Speech, Interview, Wild Scene, Comment, and Enactment) or presented in random order. There are additional reasons for forming the clusters. First, a composition might be desired based on a preference for a particular type or ordering (e.g., Wild Scene before Speech). Second, segments should be chosen from the different types for diversity, within the playout time allotment (Section 3.3). After forming clusters, the final order of segments in a news item can be determined with the following:

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

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, then the two types of mappings are 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. 7).

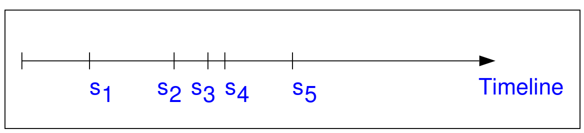

Figure 7: An Example of a One-to-One Mapping of Segments to the Timeline

- Many-to-one mapping: The start of multiple segments ($\{s_1,\ s_2,\ s_3, ...\} \rightarrow t$) maps to time t (Fig. 8).

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. Fig. 9 illustrates this type of composition.

After conforming to the structural constraints (Section 2.2.6) there is still considerable flexibility to select and order segments from the different types as we discuss next.

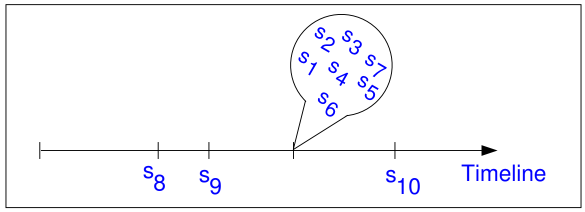

Figure 8: An Example of a Many-to-One Mapping of Segments to the Timeline

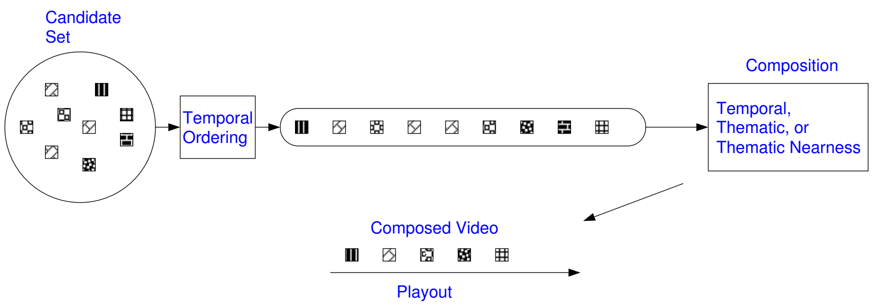

Figure 9: An Example of Period-Based Composition

3.2 Techniques for Composition of a News Item

When there are many single-presentation-type segments that are candidates for selection then we use interest-based or random selection. However, if composition is period-based or thematic ordering is required 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, thematic composition, and thematic nearness composition. This hybrid approach is illustrated in Fig. 10 in which clustering and temporal ordering are combined.

3.2.1 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. 11).

In this case the single-presentation-type segments are selected according to the rules of Table 10. The resulting composition set consists of a single segment 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.

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

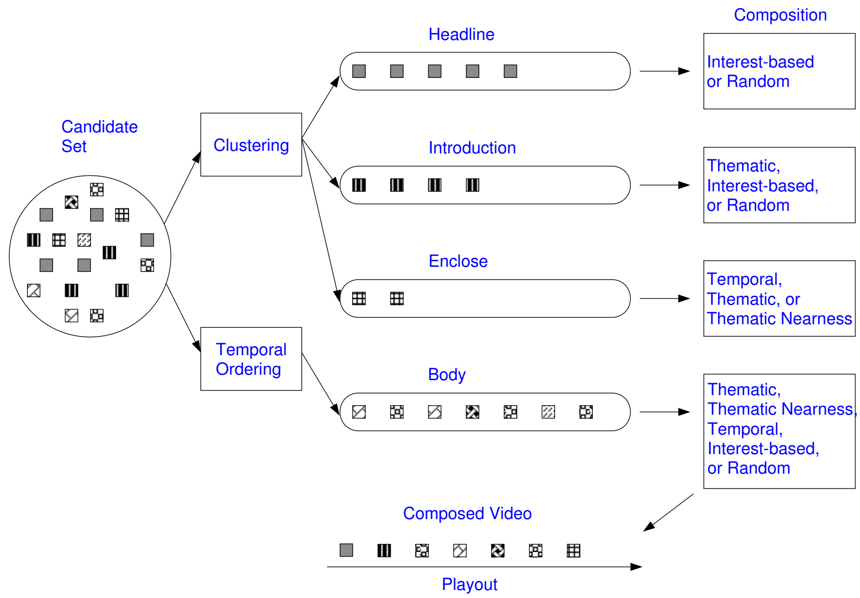

Figure 10: An Example of Hybrid Composition

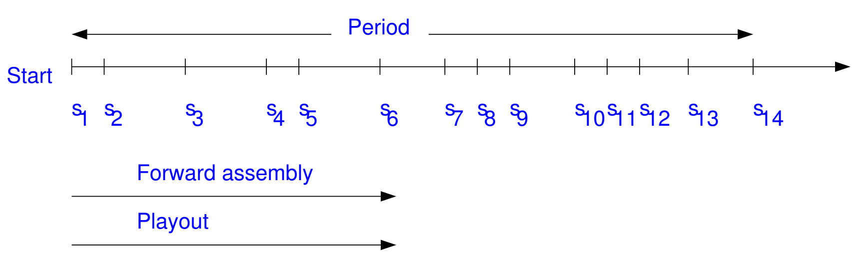

Figure 11: Forward Temporal Ordering Scheme

by temporal ordering. The segments selected to compose a news item contain information related to a story center. However, a news item develops over time and there are variations in theme due to multiple threads of the story (Section 2.2), resulting in a lower thematic continuity. An extension to this technique, thematic composition, aims to ensure that there are no large jumps in theme between consecutive segments due to these threads.

3.2.2 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 (Section 2.3) 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 is selected then an Introduction segment is selected using the concept similarity. Otherwise, if an Introduction segment is selected first, then a segment belonging to the body is selected based on concept similarity.

Initially, all segments belonging to the body are chronologically ordered and the predicate in Eq. 1 must hold for all segments. 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, τ and λ. If the similarity value of two segments is more than τ, 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 λ, then the two segments are not considered similar. When all pairs are exhausted, Eq. 2 should be valid for all consecutive segments s_i and s_j in the composition:

$$\lambda \leq \text{cosine}(\vec{W}_i, \vec{W}_j) \leq \tau \tag{2}$$

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

Creation Instance: When building a composition for a creation instance, we begin by selecting a Headline or Introduction segment according to the interest-based or random technique. If segment of type Headline is selected to start the composition, then segments 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. The next segment for the composition is selected based on concept similarity. 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.

Some segments from the groups may 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, segments are selected by evaluating concept similarities between segments along the timeline. The rules for selection of single-presentation-type segments (Table 10) and the selection of segments from clusters of multiple-presentation-types are the same as for compositions for a creation instance; however, the generated composition must be valid for the predicate of Eq. 1.

Under these conditions, the final composition will have a better thematic continuity than the previous scheme, but can also possess large forward time discontinuities and loss of information In. This is evident in the analysis of Section 4.2.

3.2.3 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 concepts). 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:

$$IS(s_i, s_j) = A \times \frac{\mathrm{cosine}(\vec{W}_i, \vec{W}_j)}{b_i - b_j}, \tag{3}$$

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) then use only cosine metric for measurement of similarity between the two segments. This type of composition will result in lower relative thematic continuity, but the occurrence of dropped segments or temporal discontinuities is reduced. This is evident in the evaluation in Section 4.3. Fig. 12 illustrates the relationships among temporal, thematic, and thematic nearness composition.

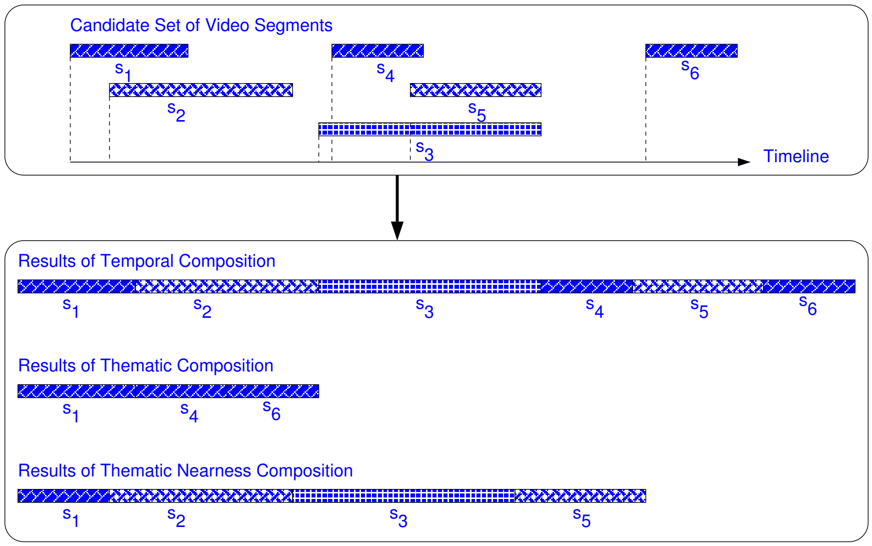

Figure 12: Relationships Among Composition Techniques

3.3 Composition Under Temporal Constraints

The aforementioned composition approaches allow 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. We base our approach 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 a single item d_c is reduced to d_{S_c}. The two cases are considered below.

3.3.1 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, can be used as filler. The following Filler Algorithm leads to the accommodation of the leftover time $(d_u - d_c)$. The algorithm takes a collection of composed sequences (e.g., news items) and the leftover time and produces composed sequences augmented by additional segments in the bodies of the compositions.

Filler Algorithm:

1 Determine the leftover time for each composition
2 For each composition in the newscast
  2.1 If a composition contains Wild Scenes then
    2.1.1 For each Wild Scene segment in the composition and a nonzero leftover time
      2.1.1.1 If the segment duration is less than or equal to the leftover time then
        2.1.1.1.1 Include the segment in the composition
        2.1.1.1.2 Decrease the leftover time by the included segment duration
    2.1.2 If the leftover time is greater than zero then
      2.1.2.1 For each Wild Scene segment
        2.1.2.1.1 If the segment is not already selected as filler then
          2.1.2.1.1.1 Select a partial segment with playout duration equal to the leftover time
  2.2 Redistribute any leftover time of the composition to the remaining compositions
3 End

Recognized problems with this approach are the fragmentation due to the introduction of incomplete segments and the favoring of later compositions that can be allocated leftover time not used by earlier ones. An alternative approach is to introduce a completely different type of filler such as advertisements.

3.3.2 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 τ so that additional segments are considered to have the same content and are eliminated from the composition. λ can be increased for a similar result. 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. Then 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 11 are used to form complete and cohesive news items. These rules dictate the time allocated for each item while preserving cohesion.

Table 11: 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 a segment of type Headline is dropped.
2.	$(s \in S_{in}) \subset S_c$	If available, a segment of type Introduction is always included in the news item.
3.	$((\exists(s \in S_b) \subset S_c) \wedge (d^r_{S_c} > 0))$. $\Rightarrow ((s \in S_e) \subset S_c)$	If there is a segment in news item belonging to body and there is time available to accommodate it then a segment belonging to Enclose will be included.
4.	$d_{S_c} = d_u/N_{ni}$	Initially, each item is allocated a duration $\frac{d_u}{N_{ni}}$.
5. 6.	$d^r_{S_c} > 0 \Rightarrow$ use excluded segments $d^r_{S_c} > 0 \Rightarrow$ use filler algorithm	Draw upon previously excluded segments. If no segment type is excluded or none of the segments from the excluded set can be adjusted then if segments of type Wild Scene exist in a news item, these segments are used to adjust the remaining time $d^r_{S_c}$ according to the filler algorithm.

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 segments from the body of the news items. We propose 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 depth of each segment type. 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 rules for this type of composition are shown in Table 11. 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 inputs 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 then
  2.1 Decrease the allocated time by the current segment duration
  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 (no composition fits the time allocation)

The example below (Table 12) 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

Table 12: Example of Creation Instance Time Adjustment

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

the result: Introduction, Speech$_1$, Speech$_2$, Wild Scene$_1$, Comment$_1$, Wild Scene$_2$, QA$_{11}$, QA$_{21}$.

Creation Period Adjustments: For the creation-period case, we attempt to incorporate segments from most of the creation period. We divide a creation period into sub-periods TP_i to differentiate days on the timeline. Fig. 13 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.

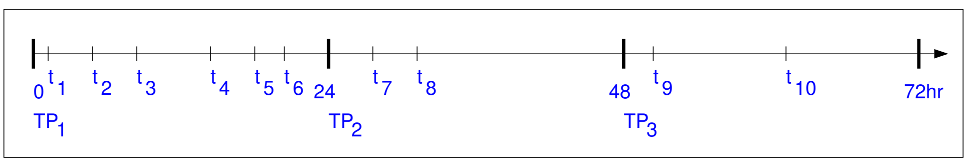

Figure 13: Dividing Periods for Temporal Constraint Composition

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. 13. 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. Once 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 11 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 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)

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 portion of the period for which we have data available.

In each case, the breadth-first and depth-second approach can be used for composition.

To validate the proposed techniques we compare resultant compositions with manually-assembled broadcast news using the the metrics defined in Section 2.2.

4 Evaluation and Analysis of the Proposed Automatic Composition Techniques

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. We evaluate the results with respect to temporal, thematic, thematic nearness, and time-limited composition techniques.

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

4.1 Temporal Ordering

Table 13 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, δ, 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.

Table 13: Evaluation of Period-Based Temporal Ordering

Composition	No. of Segments	Information (In)	Thematic Continuity	Content Progression	Temporal Continuity	Period Span Covered
1	18	1.0	0.85	0.91	1.0	1.0
2	91	1.0	0.92	0.96	1.0	1.0
3	79	1.0	0.72	0.96	1.0	1.0
4	31	1.0	1.0	0.96	1.0	1.0

4.2 Thematic Composition

By using the data from the first two compositions in Table 13 we composed two news items with a constant similarity threshold value $\tau = 1$ and different values of dissimilarity threshold

λ to study the thematic composition technique. The results for these two news items are shown in Tables 14 and 15. Note that identical results were obtained within the value ranges specified in the table for λ.

The results show as the value of λ increases the thematic continuity increases and the information (In) value decreases. The number of segments in a composition, temporal continuity, and temporal span covered, do not show distinct patterns. This is because different values of λ lead to compositions following different threads in the storyline. For lower values of λ, all threads 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 14: Evaluation of Thematic Continuity: Composition 1

S. No.	**No. of Segs**	λ	**Information (In)**	**Thematic Continuity**	**Content Progression**	**Temporal Continuity**	**Period Span Covered**
1	18	0.1 - 0.42	1.0	0.86	0.91	1.0	1.0
2	15	0.43	0.82	0.93	0.94	1.0	1.0
3	14	0.44 - 0.46	0.77	0.89	0.93	1.0	1.0
4	4	0.47 - 0.5	0.19	0.90	0.96	1.0	0.0008
5	3	0.51 - 0.52	0.14	0.93	0.95	1.0	0.0007
6	2	0.53 - 0.56	0.10	0.99	0.93	1.0	0.0007
7	2	0.57 - 0.58	0.09	1.0	1.0	1.0	0.0008
8	8	0.59	0.42	1.0	0.96	1.0	0.99
9	5	0.6	0.25	1.0	0.95	1.0	0.99
10	4	0.61	0.19	1.0	0.96	1.0	0.99
11	3	0.62	0.14	1.0	0.95	1.0	0.99
12	1	0.63 - 1.0	0.05	NA	1.0	NA	0.0

4.3 Thematic Nearness Composition

Thematic nearness composition is studied for the first two compositions of Table 13. The results are summarized in Tables 16 and 17. The data indicate that thematic continuity is usually not as high as compared to the thematic composition technique, but higher than using temporal ordering by itself. 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

Table 15: Evaluation of Thematic Continuity: Composition 2

S. No.	**No. of Segs**	λ	**Information (In)**	**Thematic Continuity**	**Content Progression**	**Temporal Continuity**	**Period Span Covered**
1	91	0.1 - 0.47	1.0	0.92	0.96	0.99	1.0
2	4	0.48 - 0.58	0.037	1.0	0.95	1.0	0.000056
3	6	0.59 - 0.61	0.06	1.0	1.0	1.0	0.004463
4	3	0.62 - 0.63	0.02	1.0	1.0	1.0	0.000047
5	2	0.64 - 0.74	0.018	1.0	1.0	1.0	0.000047
6	1	0.75 - 1.0	0.008	NA	1.0	NA	0.0

are covered in the composition, and therefore, more information is covered as well. For these compositions the normalization constant A is 50.

Table 16: Evaluation of Thematic Nearness: Composition 1

S. No.	**No. of Segs**	λ	**Information (In)**	**Thematic Continuity**	**Content Progression**	**Temporal Continuity**	**Period Span Covered**
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 17: Evaluation of Thematic Nearness: Composition 2

S. No.	**No. of Segs**	λ	**Information (In)**	**Thematic Continuity**	**Content Progression**	**Temporal Continuity**	**Period Span Covered**
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

For both the thematic and thematic nearness composition techniques, if the value of λ is very low during composition then the value of the thematic continuity remains within the reference values (Table 5). However, with increasing λ, thematic continuity increases but the value of information falls due to the smaller number of segments incorporated in a composition. As λ increases, the span coverage lacks a pattern due to inclusion of different threads.

4.4 Time-Limited Composition

Both the trivial and period-based breadth-first time-limited composition schemes are evaluated based on Composition 2 of Table 17 and a $\lambda = 0.00017$.

Trivial Scheme For this adjustment technique we include all sequential segments that into the time constraint. Table 18 shows the character of these compositions for a range of composition durations applied to the technique.

Breadth-First and Depth-Second Scheme Again using Composition 2 from Table 18, results are generated based on the technique and are shown in Table 19. Here the values

Table 18: Evaluation of Trivial Temporal Adjustment: Composition 2

S. No.	Duration (s)	No. of Segs	Information (In)	Thematic Continuity	Content Progression	Temporal Continuity	Period Span Covered
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

of TP_i are constant at 24 hours. The data indicate that span coverage is usually greater as 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 breadth-first and depth-second technique provides information over a greater span of topics.

Table 19: Evaluation of Breadth-First and Depth-Second Temporal Adjustment: Composition 2

S. No.	Duration (s)	No. of Segs	Information (In)	Thematic Continuity	Content Progression	Temporal Continuity	Period Span Covered
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

4.5 Summary of Evaluation

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.

5 Conclusion

With the onset of technologies to store and manipulate large amounts of digital video there is an opportunity to pursue techniques that reorganize existing video content to support personalization, customization, or specific application domain requirements.

In the news video domain, the opportunity is to support the repurposing of video segments corresponding to existing news stories and incoming news events. Such repurposing facilitates the production process including selection, editing, assembly, and dissemination; all with a goal of producing a cohesive narrative within time and topic constraints.

In this paper we have proposed and presented techniques to facilitate automatic composition of video and metrics for evaluation of the quality of video composition. We also showed a comparison of the results of these techniques as compared to manually-produced broadcast news video. We conclude that it is feasible to automatically compose video with quality comparable to broadcast production with respect to the defined metrics. We expect the results to be applicable to other video content domains beyond news, but cannot yet claim this extrapolation. Moreover, such automatic composition enables a wide variety of video-based applications with analogs in the text domain.

A Appendix

The following examples illustrate the metrics proposed in Section 2.2.

Concept	Segment Number										
	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 14: Example Concept Vector Weights

Information Evaluation Example: Consider the set of video segments and their concept vectors with binary weights as shown in Fig. 14. If all segments are incorporated in the composition then the centroid vector 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 In.

Temporal Continuity Example: Consider the creation time and date of the segments of Fig. 14 as shown in Table 20. Assume that the tolerated jump duration, δ, is 24 hours and weight β is 0.6 and consider the playout sequence $[s_1\,, s_2\,, s_3\,, s_4\,, s_5\,, s_6\,, s_7\,, s_8\,, s_9\,, s_{10}\,, s_{11}]$.

Table 20: 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
$s_{11} - s_9$	$1 - (2160)/7 \times 24 \times 60 = 0.78$

All jumps in this example are forward in time and less than δ in duration with the exception of the jump between s_5 and s_6. However, there are no data corresponding to this jump window so there is no penalty. Gaps in data are usually due the news item being off-air for long periods due to lack of new developments. As the e_{tc} for all consecutive pair 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 the sequence is shown in Table 20 and the mean temporal continuity of the sequence is 0.97.

Thematic Continuity Example: Consider a dissimilarity threshold λ of 0.6, similarity threshold τ 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. 14 and is shown in Table 21. The final result is a value of 0.76.

Content Progression Example: Consider the playout durations of the segments shown in Table 22. Assume a fast-change threshold ρ of 5 seconds and a slow-change threshold ϱ of

Table 21: 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
$s_5 - s_6$	$2/\sqrt{8 \times 6} = 0.28$	$0.28/0.6 = 0.46$
$s_6 - s_{10}$	$1/\sqrt{6 \times 5} = 0.18$	$0.18/0.6 = 0.3$
$s_{10} - s_{11}$	$2/\sqrt{5 \times 6} = 0.36$	$0.36/0.6 = 0.6$

150 seconds. The content progression of the sequence $[s_1, s_2, s_3, s_4, s_5, s_6, s_{10}, s_{11}]$ is shown in Table 22. The mean content progression of the sequence evaluates to 0.81.

Table 22: 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}	$150/300 = 0.5$

Period Span Coverage Example: Consider the segments summarized in Table 20. 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 $2/8 = 0.25$.

References

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

[2] G. Ahanger and T.D.C. Little, "A Language for the Composition of a Newscast," *Computer and Information Technology*, Vol. 6, 1998.

[3] 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.

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

[5] 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.

[6] G. Davenport and M. Murtaugh, "ConText Towards the Evolving Documentary," *Proc. ACM Multimedia '95*, San Francisco, November 1995.

[7] A. Lippman "Movie-Maps: An Application of the Optical Videodisc to Computer Graphics," *Computer Graphics*, July 1980, pp. 32-42.

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

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

[10] 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.

[11] 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.

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

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

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