Data Semantics for Improving Retrieval Performance of Digital News Video Systems[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. 01-04-1999

Abstract–We propose a novel four-step hybrid approach for retrieval and composition of video newscasts based on information contained in different metadata sets. In the first step, we use conventional retrieval techniques to isolate video segments from the data universe using segment metadata. In the second step, retrieved segments are clustered into potential news items using a dynamic technique sensitive to the information contained in the segments. In the third step, we apply a transitive search technique to increase the recall of the retrieval system. In the final step, we increase recall performance by identifying segments possessing creation-time relationships.

A quantitative analysis of the performance of the process on a newscast composition shows an increase in recall by 23% for the third step of the process and 48% for the fourth step, over the conventional keyword-based search technique used in the first step.

Keywords: News video composition, retrieval, content metadata, structural metadata, unstructured metadata, keyword vector, recall, precision.

1 Introduction

A challenging problem in video-based applications is achieving rapid search and retrieval of content from a large 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. Therein lies the need for "bridge" techniques that have rich semantics for representing motion-image-based concepts and content, yet are supported by fast and efficient algorithms for real-time search and retrieval. At this intersection we have been investigating techniques for video concept representation and manipulation. In particular we have sought the goal of automatic composition of news stories, or newscasts based on an archive of digital video with supporting metadata.

[1]In *The 8th IFIP 2.6 Working Conference on Database Semantics*, Rotorua, New Zealand, January 1999. This work is supported in part by the National Science Foundation under Grant No. IRI-9502702.

Figure 1: Scenes from an Example News Item

To retrieve video clips we need to process video data so that they are in clip-queryable form. We need to extract information from video clips, represent the information in a manner that can be used to process queries, and provide a mechanism for formulating queries. Presentation of discrete clips matching a query is not engaging. After retrieval, composition of these clips towards a theme (e.g., a news topic) adds value to the presentation.

The general process in automatic composition of news (or any) digital video towards a theme is based on selecting desired video data within some domain (e.g., sports), filtering redundant data, clustering similar data in sub-themes, and composing the retrieved data into a logical and thematically-correct order [1]. All of these tasks are possible if we have sufficient information about the content of the video data. Therefore, information (metadata) acquisition and techniques to match, filter, and compose video data are critical to the performance of a video composition system. The quality of data retrieved depends on the type of metadata and the matching technique used.

However, news audio and video (and associated closed-captioning) do not necessarily possess correlated concepts (Fig. 1). For example, it is common in broadcast news items that once an event is introduced, in subsequent segments the critical keywords are alluded to and not specifically mentioned (e.g., Table 1, the name Eddie Price is mentioned only in the third scene). Segments can share other keywords and can be related transitively. If a search is performed on a person's name, then all related segments are not necessarily retrieved. Similarly, related video segments can have different visuals. It is not prudent to rely on a single source of information about the segments in retrieval and composition (e.g., transcripts or content descriptions). The information tends to vary among the segments related to a news item. Therefore, we require new techniques to retrieve all the related segments or to improve the recall [16] of the video composition system.

In this paper, we propose a transitive video composition and retrieval approach that improves

Table 1: Example Transcripts of Several Segments

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 WITNESSES SAY HE HELD THE KNIFE TO HIS SON, LATER, TO HIMSELF. AND IT ALL PLAYED OUT 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 ON BACK TO YOU? I JUST ASSURED HIM THAT THE BABY WOULD BE OKAY...

Table 2: Content Metadata

Entity	Tangible object that are part of a video stream. The entities can be further sub-classified, (e.g., persons, and vehicles).
Location	Place shown in video. (e.g., place, city, and country).
Event	Center or focus of a news item.
Category	Classification of news items.

recall. That is, once a query is matched against unstructured metadata, the components retrieved are again used as queries to retrieve additional video segments with information belonging to the same news item. The recall can be further enhanced if the union of different metadata sets is used to retrieve all segments of a news item (Fig. 2). However, the union operation does not always guarantee full recall as a response to a query. This is because no segment belonging to a particular instance of a news item may be present among the segments acquired after the transitive search (data acquired from different sources or over a period of time containing data about the same news event).

This work is an outcome of our observations of generative semantics in the different forms of information associated with news video data. The information can be in the visuals or in the audio associated with the video. We also study the common bond among the segments belonging to a single news item. The composition should possess a smooth flow of information with no redundancy.

Annotated metadata are the information extracted from video data. In our previous work [3, 12] we have classified annotated metadata that are required for a newscast composition as content metadata and structural metadata. The content metadata organize unstructured information within video data (i.e., objects and interpretations within video data or across structural elements). Some of the information extracted from news video data is shown in Table 2. Information such as the objects present in visuals, the category of a news item, and the main concept (focus or center [7]) depicted by the new item are stored as metadata. The structural metadata organize linear video data for a news item into a hierarchy [2] of structural objects as shown in Table 3.

Table 3: Structural Metadata

1. Headline	Synopsis of the news event.	
2. Introduction	Anchor introduces the story.	
3. Body	Describes the existing situation.	
	a. Speech	Formal presentation of views without any interaction from a reporter.
	b. Comment	Informal interview of people at the scene in the presence of wild sound.
	c. Wild Scene	Current scenes from the location.
	d. Interview	One or more people answering formal structured questions.
	e. Enactment	Accurate scenes of situations that are already past.
4. Enclose	Contains the current closing lines.	

The development of the proposed hybrid video data retrieval technique is based the availability of segment metadata. We have explored the use of these data for the following reasons.

- 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 4).
- 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. 2. The precision of vector-based clustering improves by using multiple indexing schemes and multiple sets of metadata (annotated and unstructured). Unstructured data describe loosely organized data such as free-form text of the video transcripts.

The organization of the remainder of this paper is as follows: In Section 2 we describe existing techniques for video data retrieval. In Section 3 we discuss metadata required for query processing, classification of annotated metadata, and the proposed query processing technique. In Section 4 we

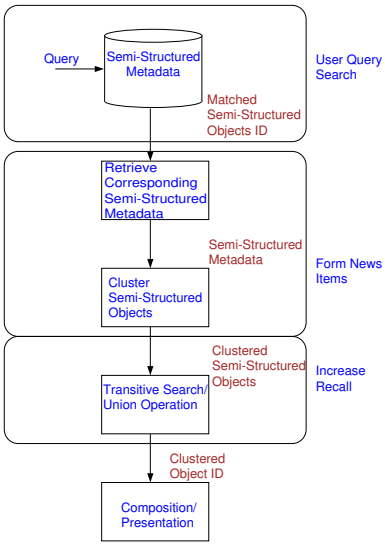

Figure 2: Process Diagram for Newscast Video Composition

present an analysis of the proposed approach. In Section 5 we present of our observations. Section 6 concludes the paper.

2 Related Work in Video Information Retrieval

A variety of approaches have been proposed for the retrieval of video data. They can be divided into annotation-metadata-based, transcript-metadata-based, and hybrid-metadata-based techniques. Each is described below.

For annotation-based techniques, manual or automatic methods are used for extraction of information contained in video data. Image processing is commonly used for information extraction in the automatic techniques. Techniques include automatic partitioning of video based on information within video data [4], extraction of camera and object motion [5, 18], and object, face, texture, visual text identification [6, 10, 11, 13, 14, 15, 17]. The metadata describing large digital video libraries can also be extracted off-line and stored in a database for fast query processing and retrieval [6].

Transcripts associated with video data can provide an additional source of metadata associated with video segments. Brown et al. [8] use transcript-metadata to deliver pre-composed news data. Wachman [19] correlates transcripts with the scripts of situation comedies. The Informedia project [20] uses a hybrid-metadata approach to extract video segments for browsing using both the visual and transcript metadata.

In the above works, keyword searching is either used to retrieve a pre-assembled news item or the segments associated with the query keywords. In this work, our objective is to search

Table 4: Symbols Used to Define the Retrieval Technique

Symbols	Descriptions
s	A video segment
S	Universe of video segments
N	Size of the universe S
R_f	A binary relationship on S for transitive search
R_u	A binary relationship on S for related segment search
tf_i	Frequency of a concept (term) i in unstructured metadata
N_i	Number of unstructured metadata components with term i
w_{1_i}	Intermediate weight assigned to a concept i for query match
w_{2_i}	Final weight assigned to a concept i for query match
w_{3_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
$d(a, b)$	The similarity distance between two sets of keywords
QS	A subset of S_q
T_c	Cluster cut-off threshold
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
S_a	A candidate set resulting from cluster TCL_i

for segments that belong to the various instances of the same event and to cover various time periods (e.g., retrieve information about Albright's trip to the Middle East). Therefore, we seek to maximize the availability of information to support the creation of a cohesive video piece. For this purpose we require, in addition to the the segments matching a query, any segments that are related via a transitive or structural relationship. In this manner, segments belonging to various instances of a news event can be merged to create a new composition. Our technique uses a four-step approach applied to both annotation-based and transcript-based (unstructured) metadata. We use a transitive search on transcripts and the union operation on structural metadata to retrieve related video segments.

3 The Proposed Four-Step Hybrid Technique

The four-step hybrid retrieval technique is based on establishing transitive relationships among segment transcripts and the use of annotated metadata. After introducing our terminology (symbols used throughout the paper are summarized in Table 4), we describe the different types of metadata and how they are used to support the four-step process.

3.1 Preliminaries

Metadata described in this paper include unstructured metadata, such as free-form text and annotation metadata. The former is used for transitive search. The latter is comprised of content metadata and structural metadata.

Unstructured Metadata and Transitivity 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. 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 we first match a query with unstructured metadata in the universe S. The results are applied as a query to retrieve additional unstructured metadata (transcripts) and associated segments, increasing the the recall of the process.

Annotated Metadata Annotated metadata consist of content and structural metadata as described in Section 1. Structural metadata exist if segments are annotated as such when they enter the segment universe S, either as video shot at a single event (e.g., a sporting event) or as decomposed segments originating from preassembled news items (as is the case for our dataset). We call such segments *siblings* if they posses either of these relationships.

A shortcoming of the aforementioned transitive search is that it may not retrieve all segments related via siblings. 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). 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 represents the candidate set of segments used as a pool to generate the final video piece (or composition set) [1]

Table 5: Sample Unstructured Metadata

.idDoc:
cnn2.txt/O193
.videoFile:
d65.mps
.textData:
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.

Hierarchical structure of related segments is stored as structural metadata that are utilized in the proposed hybrid retrieval technique (Table 3).

3.2 Segment Keyword Analysis and Weighting

We use text indexing and retrieval techniques proposed by Salton [16] and implemented in SMART [9] for indexing the unstructured metadata. To improve recall and precision we use two sets of indices, each using different keyword/term weighing. In the remainder of the paper we use s interchangeably to represent a video segment or its associated unstructured metadata. The similarity distance of a segment with a query or a segment is measured by the associated unstructured metadata.

The selection process is comprised of an initial segment weighting followed by a clustering step.

Initial Segment Weighting Initially, a vector comprised of keywords and their frequency 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. 1.

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

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 w_{1_i} to a term that is present in smaller number of segments with respect to the complete unstructured metadata. Finally, w_{1_i} is again normalized by the length of the vector (Eq. 2). Therefore, the influence of segments with longer vectors or more keywords is limited.

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

Clustering 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. 3) as compared to the initial segment weighting.

Table 6: Weight Assignment

Doc ID	Concept	Scheme 1	Scheme 2
146	barred	0.62630	4.04180
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

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

Table 6 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.

For the purpose of a query match we use the cosine similarity metric (Eq. 4) proposed by Salton. The metric measures the cosine or the measure of angle between two unstructured metadata segment vectors. The product of the length of the two segment vectors divides the numerator in the cosine metric. The longer length vectors produce small cosine similarity. In Eq. 4, n is the number of terms or concepts in the universe.

$$\text{cosine}(\vec{A}, \vec{B}) = \frac{\sum_{k=1}^{n}(a_k \times b_k)}{\sqrt{\sum_{k=1}^{n}(a_k)^2 \times \sum_{k=1}^{n}(b_k)^2}} \tag{4}$$

The proposed query processing technique is a bottom-up approach in which the search starts from the unstructured metadata. We describe the details next.

3.3 The Selection Mechanism

The four-step selection mechanism is illustrated Fig. 2. 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

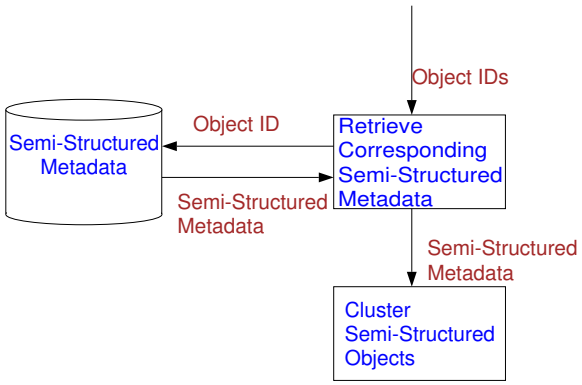

Figure 3: Process Diagram of the Clustering Process

below which we consider all the segments to be irrelevant 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 individual object in the corpus will be small. If the query has all keywords relevant to 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.

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. 3). 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:

$k \leftarrow 1$ Index on clusters

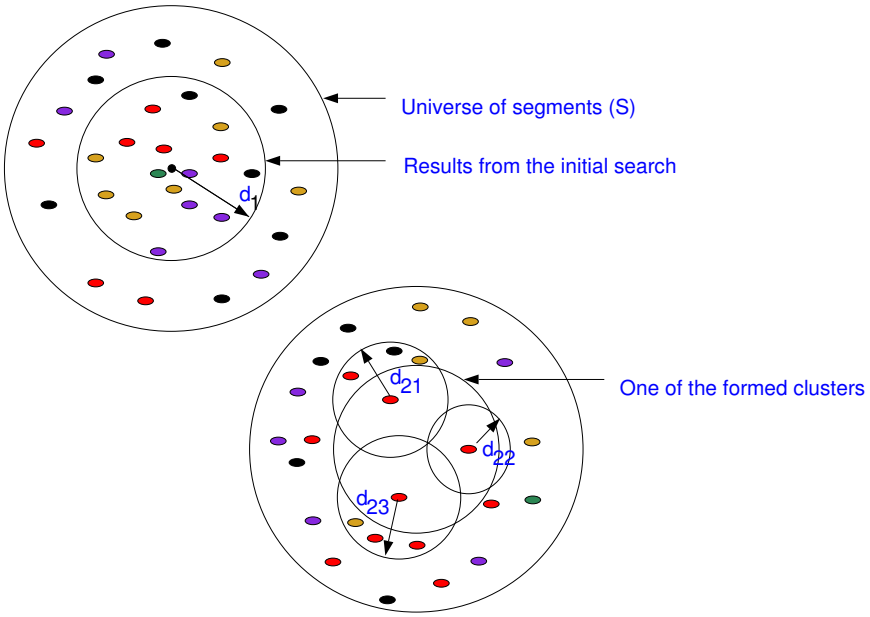

Figure 4: Similarity Measure based on the Transitive Search

For each $s_a \in QS$	Loop on segments in QS
$CL_k \leftarrow s_a$	Assign segment to the cluster
For each $s_b \in QS$	Loop on remaining segments
If $d(s_a, s_b) \geq T_c$	Segments similar to the reference?
$CL_k \leftarrow CL_k \cup \{s_b\}$	Assign segment to the cluster
$QS \leftarrow QS - CL_k$	Remove the elements from the set QS
$k \leftarrow k + 1$	Next cluster

This algorithm, although fast, is neither deterministic nor fair. A segment, once identified as similar to the reference, is removed from consideration by the next segment in the set. An alternative approach does not remove the similar element from QS but results in non-disjoint clusters of segments. We are exploring heuristic solutions that encourage many clusters while maintaining them as disjoint sets.

Transitive Retrieval We use the transitive search (Fig. 4). 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.

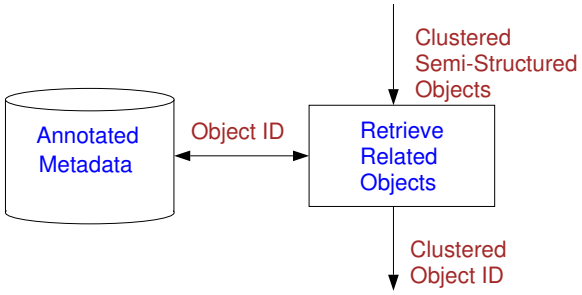

Figure 5: Process Diagram for Retrieving Related Segments

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. 4) 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 $s \in CL_i$, $d(s_t, q(s))$ 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. 5). 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 cluster 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 the segments s. Likewise, the union operation can be performed on the remaining clusters.

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.

4 Analysis of the Proposed Hybrid Technique

We evaluated the performance of our technique based on 10 hours of news video data and its 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 [1].

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 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 isreal netanyahu
isreal 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 tuker 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 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. 6. 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).

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.

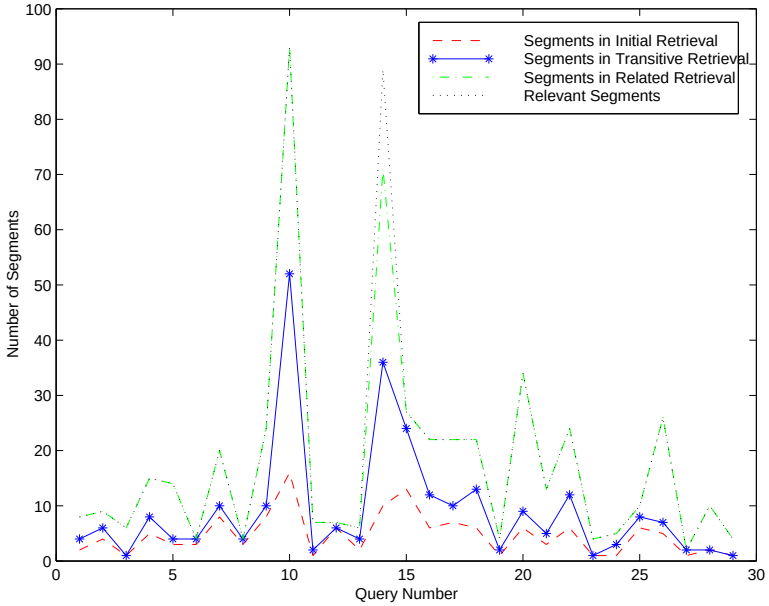

Figure 6: Summary of Performance of Different Retrieval Techniques

Table 7: 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%

As the result of the transitive search the recall of the system is increased to 48% from 25% (another level of transitive search may increase it further). 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 3.1) 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.

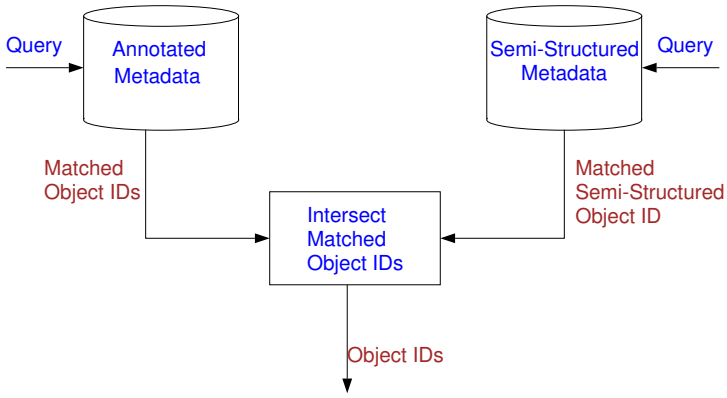

Figure 7: Process Diagram for Using Visual Metadata to Increase Precision

The results demonstrate that the combination of different retrieval techniques using different sources of metadata can achieve better recall in a news video composition system as compared to a the use of a single metadata set.

5 Observations

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.

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 (Fig. 7).

If a query retrieves a set of new items based on a date or period then access can be achieved directly from the content metadata. For the process of composition, the broader set of metadata need to be used.

6 Conclusion

In this paper we proposed a four-step hybrid retrieval technique that utilizes multiple metadata sets to isolate video information for composition. The technique relies on the availability of annotated metadata representing segment content and structure, as well as segment transcripts that are unstructured. The retrieval applies a conventional approach to identifying segments using the segment content metadata. This is followed by clustering into potential news items and then a transitive search to increase recall. Finally, creation-time relationships expand the final candidate set of video segments.

Experimental results on our data set indicate a significant increase in recall due to the transitive search and the use of the creation-time relationships. Additional work will seek a heuristic clustering algorithm that balances performance with fairness.

References

[1] G. Ahanger and T.D.C Little, "Automatic Composition Techniques for Video Production," to appear in *IEEE Trans. on Knowledge and Data Engineering*, 1998.

[2] G. Ahanger and T.D.C Little, "A Language to Support Automatic Composition of Newscasts," to appear in *Computing and Information Technology*, 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, pp. 526-533, 1997.

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

[5] A. Akutsu and Y. Tonomura, "Video Tomography; An Efficient Method for Camerawork Extraction and Motion Analysis," *Proc. ACM Multimedia '94*, San Francisco, CA, pp. 349-356, 1994.

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

[7] E. Branigan, "Narrative Schema," in *Narrative Comprehension and Film*, pp. 1-32, Rutledge, New York, 1992.

[8] 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, pages 35-43, 1995.

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

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

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

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

[13] R. Lienhart, S. Pfeiffer, and W. Effelsberg, "Video Abstracting," *Communications of the ACM*, Vol. 40, No. 12, pp. 55-62, 1997.

[14] V.E. Ogle and M. Stonebreaker, "Chabot: Retrieval from a Relational Database of Images," *Computer*, 28(2):49-56, 1995.

[15] R. Picard and T. Minka, "Vision Texture for Annotation," *Multimedia Systems*, 3(3):3-14, 1995.

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

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

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

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

[20] H. Wactlar, T. Kanade, M.A. Smith, and S.M. Stevens, "Intelligent Access to Digital Video: The Informedia Project," *Computer*, 29(5):46-52, 1996.