TOPIC-BASED INTER-VIDEO STRUCTURING OF
A LARGE-SCALE NEWS VIDEO CORPUS

Ichiro IDE, Hiroshi MO, Norio KATAYAMA, Shin’ichi SATOH

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan
TEL: +81-3-4212-2585 FAX: +81-3-3556-1916
EMAIL: {ide,mo,katayama,satoh}@nii.ac.jp

ABSTRACT

We propose a topic-based inter-video news video corpus structuring method and a visual interface to efficiently browse through the structured corpus. Such inter-video structuring was not deeply sought in previous works. The topic-based structure is analyzed by closed-caption text analysis; topic segmentation and tracking. The visual interface provides the ability to 1) search and select a topic by query terms and 2) track a topic thread interactively referring to the text analysis results. Although topic retrieval is somewhat similar to conventional video retrieval methods, the combination with topic tracking makes it remarkably easy to narrow down the results that match a user’s interest and moreover reveal underlying content-based structures, where the structure itself contains rich information.

1. INTRODUCTION

Due to the recent development of telecommunication technology, large amounts of videos have become available. Such video data contain various human activities, which could be considered as valuable cultural and social properties of the human race. From this viewpoint, news videos contain such information most densely. Nonetheless, building and analyzing a large-scale news video corpus has not been thoroughly examined until recently, due to limitation of computation power and storage size.

Motivated by such background issues, we have built an automatic news video archiving system where important topics should be tracked easily. It automatically records video image, audio, and closed-caption text, and archives them in an Oracle database. Up to now, approximately 315 hours (175GB of MPEG-1 and 1TB of MPEG-2 videos, and 17.5MB of closed-caption text data) have been archived from a Japanese daily news program.

In this paper, we introduce a topic-based news video corpus structuring method and a visual interface to efficiently access the structured corpus. The paper focuses on topic segmentation and tracking methods, as well as visual interfaces that provide the users with the ability to browse the corpus based on the topic structure.

Topic segmentation and tracking is in general a part of the “Topic Detection and Tracking (TDT) task” defined by NIST [1]. In TDT documents [2], a topic is defined as “A seminal event or activity, along with all directly related events and activities”. Nonetheless, as the term “topic” generally stands for the TDT defined event in news video analysis, we will use the term “topic” to indicate both a topic and an event in this paper.

First, structure analysis, i.e. topic segmentation and tracking methods are described in Section 2, and next the visual interface is introduced in Section 3. Finally, Section 4 summarizes the results and future works.

2. STRUCTURING A NEWS VIDEO CORPUS

2.1. Structure of news videos

In the case of news videos, image and content-based structures are roughly comparable. Thus majority of previous works on structure analysis of news videos has focused on analyzing the image-based structure to subsequently acquire the content-based structure, under the assumption that there are rules that link them. Although this approach works well to some extent, the rules are not always applicable and also depends highly on the editing and designing policy of each program. Moreover, such structure analyses are limited within a single video (intra-video structuring), and inter-video structure has not been deeply sought. Since we deal with a large-scale corpus, content-based inter-video structure analysis (i.e. topic tracking) becomes exceptionally important. Such analysis will reveal the underlying content-based structure of the entire corpus which is not simply a large volume of unrelated data, but data full of rich information in the relational structure itself that does not emerge from simple intra-video analysis. Although there are several works that deal with news video corpora of a comparable size to ours such as [3] and the Informedia News-on-Demand project [4], they do not consider inter-video structures.

Figure 1 illustrates the basic concept of the content-based intra/inter-video structures of a news video corpus. As a first step to realize such structuring, we will analyze closed-caption texts in order to enable topic-based structuring (i.e. segmentation and tracking). This is based on the assumption that text contains higher level semantics compared to information easily acquirable from image. Note that although this paper concentrates on text-based structuring, it is a starting point of a multimedia-integrated analysis. Image-based analyses such as topic tracking based on image features [5] will be integrated in future works.

2.2. Topic segmentation and tracking

2.2.1. Related works

Various approaches for topic segmentation and tracking have been proposed and evaluated in the past TDT workshops and by related research community, but since they deal with mostly En-
English and Chinese texts, they are not directly applicable to Japanese news texts. Among the few works targeting Japanese news video structuring, Takao et al. [6] proposed a method that segments news speech transcription texts, although they only consider segmentation (i.e., intra-topic structuring). Moreover, a multimedia-integrated approach to both topic segmentation and tracking has not been sought in previous works. Such an approach will be effective and essential to deal with a large-scale news video corpus.

2.2.2. Topic segmentation

Topic boundaries are detected by applying the following procedure to daily closed-caption texts:

1. Concatenate original time-stamped text lines into single sentences. Sentences are concatenated by detecting a period.

2. Apply morphological analysis to each sentence to extract noun sequences. JUMAN [7], a Japanese morphological analysis software, is used.

3. Apply semantic analysis to the noun sequences, and generate a keyword vector for each semantic class. Semantic analysis is done by a suffix-based method [8], which classifies noun sequences to 1) general, 2) personal, 3) locational/organizational, or 4) temporal. Thus, four keyword vectors are generated from a sentence: $K_s, K_p, K_l, K_t$, respectively. The vectors have keywords (noun sequences) as indices and frequencies as values. Note that this analysis method classifies only proper noun sequences (e.g., Prime Minister Koizumi) but also common noun sequences (e.g., fire fighter).

4. Set a window size $w$, and evaluate relations between $w$ preceding and succeeding vectors at each sentence boundary. The relation at the boundary between sentences $i$ and $i+1$ is defined as follows:

$$R_S,i = \frac{\sum_{m=i-w+1}^{i+w} K_S(m) \cdot \sum_{n=i+1}^{i+1+w} K_S(n)}{\sum_{m=i-w+1}^{i+w} K_S(m) \cdot \sum_{n=i+1}^{i+1+w} K_S(n)}$$

where $S = \{g, p, l, t\}$ and $i_{max}$ stands for the number of sentences in a daily closed-caption text. We set $w = 1, 2, ..., 10$ in the following experiment$^1$.

5. Evaluate the following function to detect topic boundaries:

$$R(i) = \sum_{S = \{g, p, l, t\}} a_S \max_{w} R_{S, w}(i)$$

First, the maximum of $R_{S, w}(i)$ along the $w$ axis is taken. According to a preliminary observation, although most boundaries were correctly detected regardless to the window size, there was a large number of over-segmentation. The over-segmentation had the following tendencies:

- Small $w$: Tends to over-segment long topics
- Large $w$: Tends to over-segment short topics

Thus, taking the maximum should compensate for over-segmentation at various window sizes.

Next, weighted sum of relations evaluated in separate semantic attributes is defined to evaluate the overall relation. This approach is taken under the assumption that especially in news texts, certain semantic attributes should be more important than others when considering topic segmentation. Multiple linear regression analysis was applied to manually segmented training data (consists of 39 daily closed-caption texts, with 384 boundaries), which resulted in obtaining the following weights:

$$(a_g, a_p, a_l, a_t) = (0.23, 0.21, 0.48, 0.08)$$

The obtained weights indicate that temporal noun sequences are not important in segmentation, and that locational/organizational noun sequences are especially important, which matches with our intuition.

Finally, if $R(i)$ does not exceed a certain threshold $\theta_{seg}$, the boundary between sentences $i$ and $i+1$ is judged as a topic boundary.

6. Create a keyword vector $\vec{K}_S$ for each detected topic, and re-evaluate the relations between adjoining topics $i$ and $j := i+1$ by the following function to concatenate over-segmented topics:

$$R(i, j) = \sum_{S = \{g, p, l, t\}} a_S \frac{|\vec{K}_S(i) \cdot \vec{K}_S(j)|}{|\vec{K}_S(i)| \cdot |\vec{K}_S(j)|}$$

As for $a_S$, the same weights as defined in (1) were used. If $R(i, j)$ does not exceed a certain threshold $\theta_{cat}$, the adjoining topics are concatenated. This process is continued until no more concatenation occurs. After the concatenation, topics with only one sentence are excluded since they tend to be either noisy or relatively less important in a large-scale corpus.

Figure 2 shows the recall-precision curb of topic boundary detection derived from applying the proposed method to a test data set (14 days) independent from the training data set. Superiority of employing the weighted segmentation is shown by comparing it with the unweighted segmentation. $\theta_{seg}$ was defined as 0.17

$^1$The range was set reflecting the fact that 94% of the topics in a manually segmented data ranged from 1 to 10 sentences per topic.
are tracked. Figure 3 shows an example of actual topic threads involving topic #1 on March 1, 2002, extracted from the corpus. Only linked topics are considered as related, and topics placed in parallel are considered not related among themselves. The threads are structured so that always topics linked to the left should be preceding, and to the right should be succeeding ones.

3. VISUALIZING THE TOPIC-BASED STRUCTURE

We implemented a topic browsing interface, namely the “Topic Browser” to visualize the topic-based structure analyzed in Section 2. It consists of two interfaces: the “Topic Finder” and the “Topic Tracker”.

The “Topic Finder” (Figure 4) is a portal to the topic browsing interface. First, a user types in a query term. Then the interface returns topics that contain the query term in chronological order. Each topic segment is represented by a thumbnail image (the first video frame of the topic segment) and an excerpt of a closed-caption text. Users browse through them and select the most relevant one to their interests and set them as initial topics for the tracking process. The right side of the browser displays the video and the closed-caption text corresponding to the topic in focus. Figure 4 shows an example of the result of a query “Bin Laden”.

Next, the “Topic Tracker” (Figure 5) is an interface to track up and down a topic thread interactively. Although the initial topic should be selected through the “Topic Finder”, the consecutive tracking is done solely within this interface. The interface displays relevant topic threads in chronological order, separated in two categories: preceding topics and succeeding topics, reflecting the topic-based structure as exemplified in Figure 3. Here, the terms “preceding” and “succeeding” represent the chronological relations with the topic in focus. The user could either track anterior or posterior sequence of topics by selecting one of the presented threads, and setting it as the next selected topic. Such interactive tracking goes on topic after topic. Figure 5 shows the list of topic threads starting with the topic selected in Figure 4. Each thread represents a different subject related to the terrorism on September 11th and its aftermath.

The “Topic Finder” may seem somewhat similar to conventional keyword-based news video retrieval, but the combination with the “Topic Tracker” narrows down the result according to the users’ interests and intentions which is important when dealing with a large-scale corpus. On the other hand, while narrowing down the results, the tracking could also be considered as a query expansion process. Having the two characteristics, the tracking process provides a user with a topic thread that matches their interests and intentions to the maximal extent. Moreover, it reveals chronological transition, divergence, and merger of a topic thread, which will be effective for thorough understanding. We found the above mentioned features very effective and informative after trying to track down several topics of interest.

Similar approaches are taken in [9], although they fix the brows-
4. CONCLUSION

In this paper, we proposed a topic-based news video structuring method as a first step to reveal underlying structures within a large-scale news video corpus. First, methods to segment and track topics by closed-caption text analysis were described and evaluated. Next, visualization of the topic-based structure reflecting the segmentation and the tracking was introduced. Although detailed evaluation is yet to be done, the visualized interface showed good browsing ability for users to retrieve and track a topic thread of interest. We will further investigate on achieving better topic segmentation quality by referring to image-based video structures. Topic tracking will also be improved by referring to graphically identical shots [5]. The tracking will be improved by dynamically adjusting the weights used in Equation (2), depending on the user’s initial query terms and tracking history in a relevance feedback manner. Such adjustment should provide a user with related topics reflecting his/her intention.

5. REFERENCES