WA-TV: WEBIFYING AND AUGMENTING BROADCAST CONTENT FOR NEXT-GENERATION STORAGE TV

Hisashi Miyamori[1], *Qiang Ma*[1], *Katsumi Tanaka*[1,2]

[1]National Institute of Information and Communications Technology (NICT)
[2]Kyoto University

ABSTRACT

A method is proposed for viewing broadcast content that converts TV programs into Web content and integrates the results with complementary information retrieved using the Internet. Converting the programs into Web pages enables the programs to be skimmed over to get an overview and for particular scenes to be easily explored. Integrating complementary information enables the programs to be viewed efficiently with value-added content. An intuitive, user-friendly browsing interface enables the user to easily changing the level of detail displayed for the integrated information by zooming. Preliminary testing of a prototype system for next-generation storage TV, "WA-TV", validated the approach taken by the proposed method.

1. INTRODUCTION

The recent introduction of hard-disk (HD) recorders for home use has greatly increased the amount of TV programming that can be recorded. The latest HD recorders have a capacity of 600GB, enabling the recording of more than 1070 hours at a certain quality. Since users do not have an unlimited amount of time to view such a great amount of content, there is a great need for functions that enable users to efficiently explore for particular video segments from a huge amount of recorded data, that present an overview of the content in a compact form, and that can provide a digest of the content in a limited amount of time.

In addition, although television programs are created by professional programmers, so that they have excellent quality and are extremely realistic in general, they suffer from time restrictions (on-air time) and an obligation to accommodate popular tastes. That is, programs must serve the public interest and appeal to a mass market. They are thus limited in the amount of detail and scope of information they can provide. The need to access information related to TV programs that provides more detail or presents multiple perspectives will become more and more important as people's lifestyles become more diversified. Thus, there is a great need for functions that can efficiently access and present related information not provided by television programs.

While various viewing methods have been proposed for easy scene exploration from a multitude of videos or for gaining a quick understanding of the content [1][2][3]. They are based on simply summarizing the content. Few are based on augmenting the information presented or consolidating the accessible information by associating the related information on the Web.

A method is proposed for viewing broadcast content that converts TV programs into Web content and integrates the results with complementary information retrieval retrieved using the Internet. Conventionally, Web pages are browsed on a PC (active browsing) while TV programs are being watched on TV (passive watching). By converting the programs into Web pages, our method enables programs to be viewed using active browsing. That is, the user can skim over the programs to get an overview of them and can easily explore for particular scenes. By integrating complementary information, it enables programs to be viewed efficiently with value-added content. In addition, an intuitive, user-friendly browsing interface enables the user to changing the level of detail displayed for the integrated information by zooming. Preliminary testing of a prototype system for next-generation storage TV, "WA-TV", validated the approach taken by the proposed method.

The rest of the paper is organized as follows. In Section 2, an overview and the processing steps of WA-TV are presented. Results of the preliminary experiments are presented and discussed in Section 3. The paper is summarized in Section 4.

2. OVERVIEW OF WA-TV

Our WA-TV (**W**ebifying and **A**ugmenting TV) provides a new method for viewing broadcast TV programs and is a prototype system for next-generation storage TV. As mentioned above, TV programs are converted into Web content and integrated with complementary information retrieved using the Internet. WA-TV has three major characteristics:

0-7803-9332-5/05/$20.00 ©2005 IEEE

- TV programs are converted from broadcast media to Web content, i.e., webification.
- The Web content is augmented with complementary information.
- The browsing interface uses zooming operation.

The webification enhances the browsability of the TV program and provides a basis for augmenting information. The retrieval of complementary information enables viewing of information not provided by the original program. The zoom-operation interface enables the user to control the level of detail displayed, providing intuitive and easy-to-understand browsing.

The processing steps of the proposed system are shown in Fig.1. First, a TV program is recorded and hierarchically segmented into topics using information in the program's closed captions. The segmented closed captions and corresponding scenes are grouped into pairs and then presented in the form of a storyboard on the screen. The retrieved complementary information is integrated at the corresponding positions in the storyboard. The display of the integrated information is controlled using zooming operation. The sizes of the displayed images of the segmented scenes can be changed smoothly, and the storyboard can be switched from one to another with a different level of detail. Users can thus seamlessly move back and forth between storyboard screens with different levels of detail and the normal playback screen, enabling them to easily explore for specific scenes. Moreover, hyperlinks to the related information are integrated in each storyboard, so users can efficiently access the related information at different levels of detail or from different perspectives depending on the situation. The processing steps are explained in the following subsections.

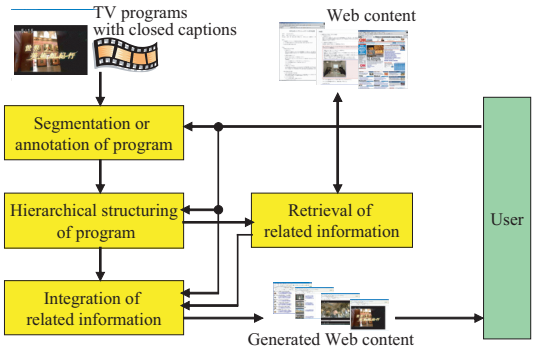

Fig. 1. Processing steps in WA-TV

2.1. Segmentation and Hierarchical Structuring

As mentioned above, hierarchical segmentation uses the information in the closed captions.

Closed captions are generally composed of a set of (t, s), where t denotes time and s denotes a single sentence representing the speech content.

There are two main approaches to topic segmentation using closed captions. The first approach is based on learning, which is problematic because it limits domain, making it difficult to apply to general topics. Moreover, it needs sufficient training data. The other approach is based on statistical computation, such as calculation of word distribution[4]. This is the approach we use. It does not need training data and can be applied to any topic domain.

Topic segmentation using word distribution is based on three assumptions.

- A topic can be identified based on word distribution.
- Different topics have different word distributions and are statistically independent of each other.
- The words in a topic are statistically independent.

Let $W = w_1 w_2 ... w_n$ be a text consisting of n words, and let $S = S_1 S_2 ... S_m$ be a segmentation of W consisting of m segments. The probability of the segmentation S is defined by $Pr(S|W) = Pr(W|S)Pr(S)/Pr(W)$. The most likely segmentation, $\hat{S}$, is given by $\hat{S} = argmax_S Pr(W|S)Pr(S)$ because $Pr(W)$ is a constant for a given text, W. $Pr(W|S)$ and $P(S)$ can be represented by the number of words in a temporary segment, the number of different words in W, etc.[4]. We define the cost of segmentation S as $C(S) \equiv -logPr(W|S)Pr(S)$ and find the $\hat{S}$ that minimizes $C(S)$. The most likely segmentation $\hat{S}$ can be obtained from a graph in which the nodes and edges are composed of segmentation gaps and segments, respectively, and from the identification of the path minimizing $C(S)$ using dynamic programming algorithm.

Segmentation into topics is done by applying the above-mentioned segmentation to $\{s_1, s_2, \ldots, s_l\}$, the set of closed caption sentences (primary segmentation). Similarly, segmentation into subtopics is obtained by applying it to $\{s_1, s_2, \ldots, s_l\}$, the set of closed caption sentences segmented into topics (secondary segmentation). The video data are then segmented based on the times for the segmented closed captions.

The original TV program is now segmented into data at three different levels of detail, each having units of topics, subtopics, and sentences. The result is structured hierarchically, as shown in Fig.2.

2.2. Complementary IR and Integration

Complementary information retrieval[5] is a method of retrieving information in greater detail or from different perspectives, which cannot be done using a conventional similarity search. It works by extracting data called topic structures from the information in closed captions, creating sev-

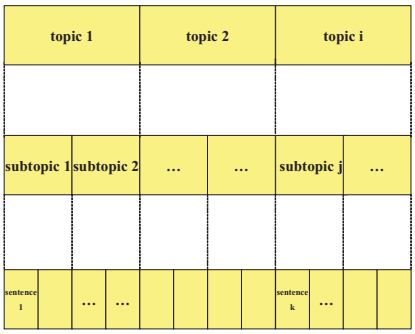

Fig. 2. Hierarchical structure of segmented data

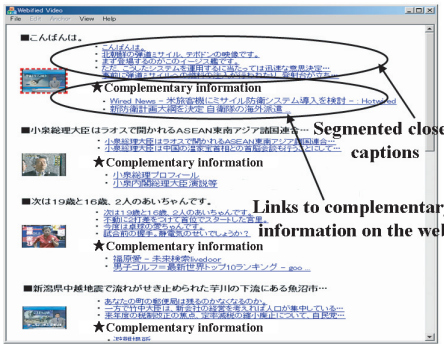

Fig. 3. Example screen displayed on WA-TV

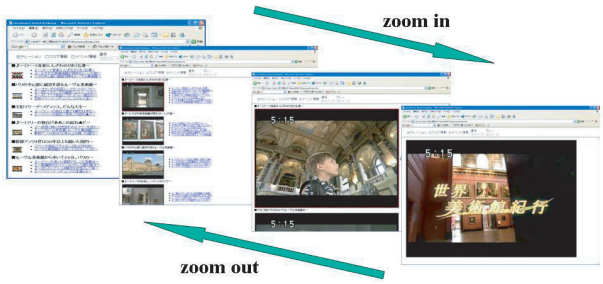

Fig. 4. Transformation of screen appearance by WA-TV

eral structured queries based on these structures, and performing Web searches using a search engine[5].

A topic structure is composed of a pair of terms; one is the subject, and the other is content. The subject term is dominant. Our method selects as the subject terms the keywords that appear most frequently in the closed captions and that have the strongest co-occurrence relationships with other keywords. It selects as the content terms those terms that have strong co-occurrence relationships with the subject terms. In other words, the subject terms are the keywords that play the role of a title while the content terms play the role of the body or content description.

By distinguishing terms as titles words or content words, we can describe structured queries and search for information from various perspectives. The structured queries can be designed to search for greater detail about the content titles or bodies or to search for detail about the content titles or bodies from different perspectives. Therefore, this method can retrieve information that is similar but not the same, such as information that is similar in terms of subject but not in terms of content, as well information that is similar in terms of both subject and content.

2.3. Zooming-based Browsing Interface

The structured program data (three layers: topics, subtopics, and sentences) is integrated with the retrieved related information into Web content for display.

An example screen displayed on WA-TV is shown in Fig. 3. The segmented caption texts and videos are displayed vertically in the form of a storyboard. Hyperlinks to the complementary information are located below the caption texts, enabling users to access more detailed or broader information than provided by the original program.

The transformation of the screen appearance is illustrated in Fig. 4. The zooming feature can be used to smoothly change the size of the thumbnails as well as to switch from one storyboard to another with a different level of detail. For example, zooming-in operation smoothly changes the size of the thumbnails on a storyboard representing topic 1, 2, ..., i in Fig. 2 and, when their size reaches a certain level, it switches to another storyboard including subtopic 1, 2, ..., j. Further zooming-in operation smoothly change the size of the thumbnails on the storyboard and when their size reaches another certain level, it switches to the other storyboard including sentences. Further zooming-in operation finally switches it to the normal playback screen. The zooming-out operation has totally opposite effect. As a result, users can seamlessly move back and forth between storyboard screens showing different levels of detail and the normal playback screen, enabling them to easily explore for specific scenes.

3. EXPERIMENTS AND DISCUSSION

We conducted experiments to validate our approach to hierarchical topic segmentation using closed captions. Various kinds of TV programs (news, documentary, variety, drama, sports, etc.) were hierarchically segmented into topics and subtopics (primary and secondary segmentation). As for the news programs, the precision and recall of the primary segmentation were about 0.6 and 0.4 respectively, and those of the second segmentation were about 0.5 and 0.4 respectively. Note that the results were calculated based on the strict matches of the segmented positions. Subjectively

speaking, the primary segmented results were almost on the topic boundaries for the news programs. A typical problem was the division of a single topic into more than two segments, which occurred when the topic was composed of more than 20 sentences, such as a lengthy news story. This should not be a major problem because a lengthy news story can be easily understood even if it is divided into a few segments. We also determined that the segmentation method needs to be enhanced so that it does not depend only on word distribution, in case of documentary, variety, drama, and sports, because topic boundaries tend to be ambiguous. Segmentation based on words, audio, video, etc. indicating context changes may be necessary.

The precision of complementary information retrieval was evaluated using the following types of structured queries, which were made using Google's application programmer interface.

1. queries searching for more detail
    - more detail about the subject
    - more detail about the content
2. queries searching for broader information
    - broader information about the subject
    - broader information about the content

The appropriateness of the retrieved complementary information was examined using news content, which could generally be segmented into topics without problem. Based on complementary web pages selected by two users, the precision ratios of queries searching for more detail and for information from multiple perspectives were computed to about 0.6 and 0.7, respectively. Evaluation tests using more participants will be conducted to confirm these results.

Simple experiments were conducted to evaluate the zoom-operation interface. Most of the participants (8 out of 11) found the ability to search for scenes by looking through a list of closed captions and thumbnails on the storyboard "useful" compared to the conventional interface based on fast-forwarding and rewinding. They also found the ability to control the different levels of detail "intuitive". Evaluation tests using more participants will be conducted to better evaluate usability, understandability, etc.

An important advantage of WA-TV is that it enables active browsing of TV programs, which are conventionally viewed by passive watching, by converting them into Web content. WA-TV enables active browsing by integrating hyperlinks at various positions in the program with ones to external related information with greater detail or from different perspectives.

Conversion of TV programs into Web content also enables the integration of information for multiple programs. For example, a group of programs recorded on the same date can be summarized in a list, or a group of programs having the same title and/or the same topic can be summarized in a list in chronological order. Such summaries can be displayed using video and/or audio, depending on the level of detail. Various types of summaries can be prepared in the framework of the proposed method.

4. CONCLUSION

We have proposed a new method for viewing broadcast content that converts TV programs into Web content and integrates the results with complementary information retrieved using the Internet. We also described WA-TV, a prototype system for next-generation storage TV. WA-TV segments and structures a TV program into different levels of detail, and generates hyperlinks to various positions in the program. It also retrieves information that complements the current topic using topic structures and structured queries obtained using the information in the program's closed captions. The webified program content is integrated with the retrieved complementary information. It can be viewed efficiently at different levels of detail using the zooming feature. Preliminary experiments showed that the proposed method simplifies scene exploration, and facilitates access to information not provided by the program in detail or from various perspectives.

We plan to improve our prototype by developing real-time webifying transformation, real-time browsing, and a better user-interface. We should be able to demonstrate that different styles of TV program viewing can be achieved by making WA-TV implementation easier and by conducting evaluation experiments by a larger number of participants.

5. REFERENCES

[1] Christel, M.G., Huang, C., "Enhanced access to digital video through visually rich interfaces," *ICME, MD-L5.1*, 2003.

[2] Uchihashi, S., Foote, J., Girgensohn, A., Boreczky, J., "Video Manga: generating semantically meaningful video summaries," *Proc. ACM Multimedia 99*, 1999.

[3] Sumiya, K., Munisamy, M., Tanaka, K., "TV2Web: generating and browsing web with multiple LOD from video streams and their metadata," *ICKS2004, pp.158-167*, 2004.

[4] Utiyama, M., Isahara, H., "A Statistical Model for Domain-Independent Text Segmentation," *ACL/EACL-2001, pp. 491-498*, 2001.

[5] Qiang Ma, Akiyo Nadamoto, and Katsumi Tanaka, "Complementary Information Retrieval for Cross-Media News Contents," *Proc. of ACM MMDB, pp.45-54*, 2004.