TV2Web: Generating and Browsing Web with Multiple LOD from Video Streams and their Metadata

Kazutoshi Sumiya
Kyoto University
Yoshida-Honmachi, Sakyo,
Kyoto 606-8501, Japan
sumiya@i.kyoto-u.ac.jp

Mahendren Munisamy
Kyoto University
Yoshida-Honmachi, Sakyo,
Kyoto 606-8501, Japan
mahen@dl.kuis.kyoto-u.ac.jp

Katsumi Tanaka
Kyoto University
Yoshida-Honmachi, Sakyo,
Kyoto 606-8501, Japan
ktanaka@i.kyoto-u.ac.jp

ABSTRACT

We propose a method of automatically constructing Web content from video streams with metadata that we call *TV2Web*. The Web content includes thumbnails of video units and caption data generated from metadata. Users can watch TV on a normal Web browser. They can also manipulate Web content with *zooming metaphors* to seamlessly alter the level of detail (LOD) of the content being viewed. They can search for favorite scenes faster than with analog video equipment, and experience a new *cross-media* environment. We also developed a prototype of the TV2Web system and discuss its implementation.

Categories and Subject Descriptors

H.5.2 [**User Interface**]: Windowing Systems; H.5.2 [**User Interface**]: Prototyping; I.7.m [**Document and Text Processing**]: Miscellaneous

General Terms

Design, Documentation

Keywords

video stream, metadata, level of detail, generation of Web content, Web browserfrom Video Streams and their Metadat

1. INTRODUCTION

Rapid progress in broadband and digital television technologies has made it possible to quickly provide vast amounts of information to both Internet users and television audiences[1, 2]. The fusion of Internet and digital television technologies is a problem that needs to be addressed. There are search engine technologies and directory services that enable us to easily obtain information from Web content. Digital television technologies, on the other hand, have enabled a dramatic increase in the storage capacity of digital VCR and DVD players. However, there are no effective technologies for television audiences to search for their favorite programs and other recorded information from their potentially large archives.

In this paper, we propose a method called *TV2Web*, which enables users to view video streams with the corresponding metadata, such as closed captioning, by automatically transforming the stream to Web content. The Web content includes thumbnails of the video units and captioning data generated from the metadata.

WWW2004, May 17–22, 2004, New York, NY USA.
ACM 1-58113-912-8/04/0005.

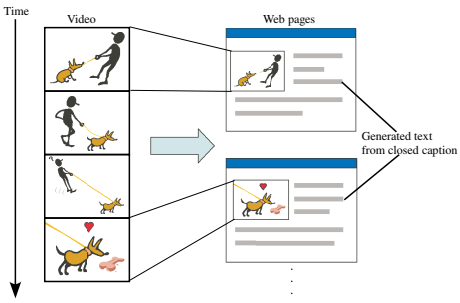

Figure 1: Basic Concept of TV2Web

Audiences can browse video and navigate content by scrolling and clicking on anchors as currently is done with ordinary web navigation. Furthermore, they can manipulate Web content through zooming metaphors to seamlessly alter the level of detail (LOD) in the content being viewed.

Figure 1 outlines the basic concept behind the TV2Web. The system extracts still images and time-code information from an original video stream and its metadata. As several topics are included in the closed captioning in the video stream, it is possible to effectively detect the divisions in the scene. Ma and Tanaka recently proposed a topic segmentation procedure that used the closed captioning of a video stream [3]. The basic idea behind the procedure called *topic segmentation for stream text* was that if the rate of keyword-pairs with high undirected coocurrence rates (pre-computed in topic corpus) among all keywords-pairs within some closed captions is high, these captions belong to one topic. We adapted the procedure to TV2Web to detect semantic scenes.

Corresponding video units are extracted through detected scenes and their time codes. The system generates Web content from the thumbnails of the video units and the text generated from captioning data. Users can interact with Web content by seamlessly switching different levels of detail on pages and by selecting a video unit. We call these interactivities *zooming* and *focusing* (Figure. 2). The levels are dependent on the length of the video units displayed on the Web page. The length of the video units are represented by the sizes of the thumbnails. Intuitively, larger thumbnails have longer video units. We developed a TV2Web prototype system, which is based on Dynamic HTML using JavaScript and HTML+TIME 2.0 to control the video thumbnails and text. TV2Web is a framework that can provide the following functions: (1)transformation of video streams as time-lined media into two-dimensional space media, (2)generation of multiple-LOD Web pages, and (3)efficient search mechanism based on seamless switching of multiple-LOD pages.

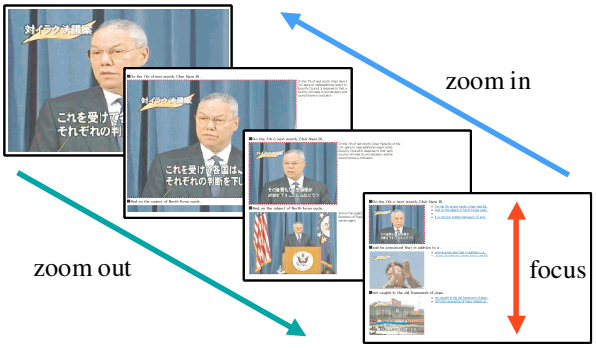

Figure 2: Seamless Switching of Multiple-LOD Pages

2. AUTOMATIC WEB CONTENT GENERATION

2.1 Topic Segmentation

We adapted the topic segmentation procedure for TV2Web by closed captioning video streams [3]. The basic idea behind the procedure called *topic segmentation for stream text* is that if the rate of keyword-pairs with high undirected cooocurrence rates (precomputed in topic corpus) among all keywords-pairs within some closed captions is high, these captions belong to one unit.

The procedure is used to compute the coocurrence of keywords belonging to adjoining closed captions. If such coocurrences are weak, the adjoining closed captions belong to different units. Hence, some portions are merged into semantic scenes. We adapted the procedure to TV2Web for detecting semantic scenes.

2.2 Transformation into Web Content

Web content is constructed of several elements, i.e., title, thumbnails, and text. We describe a method of generating Web content from a video stream with metadata. The process can be divided into two: (1) generation of metadata described by XML, and (2) dynamic generation of Web content on the fly, according to the user's interactions on the browser. Metadata is automatically generated from closed captioning data and time codes and is described by XML.

The text attached to each thumbnail is generated as follows. The range that can display the text area is calculated when the thumbnail level is changed. Intuitively, the larger the thumbnail, the smaller the range. Each thumbnail has a title text generated from the corresponding metadata. Although this is uniquely determined for the thumbnail, it is difficult to extract sentences. That is because a thumbnail has several video units, which are constructed from several scenes. In this case, there would be several candidates for the title. The size of a thumbnail is calculated by zooming in and/or out. This is continuously changed. If the size exceeds the threshold, the status of the thumbnail changes.

2.3 View-Oriented LOD Control

TV2Web provides a multiple LOD control mechanism for viewing generated Web content. The LOD control mechanism handles several elements, i.e., length of video units, sizes of thumbnails, and the level of detail in text for video units. Figure 3 outlines the relationship between them. The boxes on the left represent screens and thumbnails of video units. The thickness of rectangle represents the length of the video unit.

The rectangles on the right represent semantic scenes and the segmentation structure. The size of the thumbnail and the length of the video unit are proportional on the screen, i.e., the larger thumbnail, the longer the video unit. Semantic scenes are merged according to the number of thumbnails. For example, when there are two thumbnails, the semantic scenes are divided into two units. Both units are adjusted to be as equal as possible.

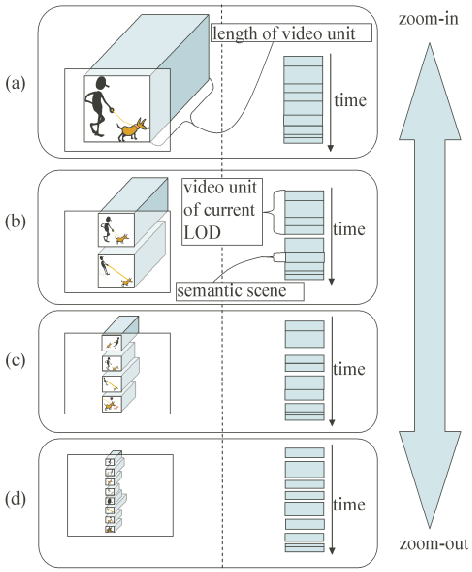

Figure 3: Size of Thumbnail and Length of Video Unit

Users can initially watch videos at full size on the display. If they do not have to do operations, they can watch the end of the video (see (a) in Figure 3). If they zoom out, the video will be divided into two smaller thumbnails (b). Furthermore, if they zoom out, the video will be divided into much smaller thumbnails (c).

2.4 Browsing Mechanism

TV2Web provides three kinds of browsing functions: *zooming in, zooming out, and focusing*. Of course, the ordinary means of navigating Web content, i.e., by clicking and scrolling, can be used to browse and view video streams. The size of the thumbnail is dynamically calculated when zooming out or in. To change video units, the size of thresholds for each video unit are prepared beforehand. If the size exceeds the threshold, the video unit is changed to another level.

3. CONCLUSIONS

We discussed a way of automatically constructing Web content from videos with metadata in this paper. The Web content could be viewed with the thumbnails of the video units and the caption data generated from the metadata on normal Web browsers. Users could navigate the content with ordinary Web navigation. We introduced zooming functions to seamlessly alter the LOD of the content being viewed as well as the viewing-oriented LOD control mechanism to dynamically generate adequate text during browsing.

4. REFERENCES

[1] TV-Anytime Forum (available from http:www.tv-anytime.org).

[2] Katsumi Tanaka, Kazutoshi Sumiya, Akiyo Nadamoto, and Qiang Ma, Broadcasting and Databases, Nontraditional Database Systems, Taylor and Francis, pp. 47–62, 2002.

[3] Qiang Ma and and Katsumi Tanaka, WebTelop: Dynamic TV-Content augmentation by Using Web Pages, Proc. of IEEE International Conference on Multimedia and Expo (ICME2003), Vol.2, pp. 173–176, 2003.