Automated Video Segmentation for Lecture Videos: A Linguistics-Based Approach

Ming Lin, University of Arizona, USA
Michael Chau, The University of Hong Kong, Hong Kong
Jinwei Cao, University of Arizona, USA
Jay F. Nunamaker Jr., University of Arizona, USA

ABSTRACT

Video, a rich information source, is commonly used for capturing and sharing knowledge in learning systems. However, the unstructured and linear features of video introduce difficulties for end users in accessing the knowledge captured in videos. To extract the knowledge structures hidden in a lengthy, multi-topic lecture video and thus make it easily accessible, we need to first segment the video into shorter clips by topic. Because of the high cost of manual segmentation, automated segmentation is highly desired. However, current automated video segmentation methods mainly rely on scene and shot change detection, which are not suitable for lecture videos with few scene/shot changes and unclear topic boundaries. In this article we investigate a new video segmentation approach with high performance on this special type of video: lecture videos. This approach uses natural language processing techniques such as noun phrases extraction, and utilizes lexical knowledge sources such as WordNet. Multiple linguistic-based segmentation features are used, including content-based features such as noun phrases and discourse-based features such as cue phrases. Our evaluation results indicate that the noun phrases feature is salient.

Keywords: computational linguistics; lecture video; multimedia application; video segmentation; virtual learning

INTRODUCTION

The quick development of technologies in the storage, distribution, and production of multimedia has created new sources of knowledge. Among these new knowledge sources, video is extremely useful for knowledge sharing and learning because of its great capability of carrying and transmitting “rich” information (Daft & Lengel, 1986). Nowadays videotaped lectures are more and more commonly provided in computer-based training systems, and they can create a virtual learning environment that simulates the real classroom learning environment. However, people often have difficulties in finding specific pieces of knowledge in video because of
its unstructured and linear features. For instance, when students want to review a certain part of a videotaped lecture, they have to look through almost the entire video or even play back and forth several times to locate the right spot.

Multimedia information retrieval technologies, such as video search engines or video browsers, try to address this problem by analyzing, transforming, indexing, and presenting the knowledge captured in videos in a structured way. For instance, in online courses provided by Stanford University (http://scpd.stanford.edu/scpd/students/onlineClass.htm), a video of an instructor is synchronized with his/her PowerPoint (PPT) slides. Students can move forward or backward to a certain segment of the video by choosing the slide associated with that segment (see Figure 1). The similar but improved design was implemented in two multimedia-based learning systems that we developed before: the Learning By Asking (LBA) system (Zhang, 2002), and its extension, the Agent99 Trainer system (see Figure 2) (Lin, Cao, Nunamaker, & Burgoon, 2003). In both training systems, each lecture video is manually segmented into short clips, and each clip is synchronized with a PPT slide as well as a text transcript of the speech in the clip. The clips are also indexed based on these text transcripts. Students can select a specific clip in the lecture by browsing a list of topics of the lecture or searching with keywords or questions. An experiment and a usability test have shown that students thought that such structured and synchronized multimedia contents, as well as the self-based learner control enabled by the list of topics, are helpful. The resulting training system is as effective as traditional classroom training (Lin et al., 2003).

However, to realize such a structured video lecture, there must be a critical pre-processing step: video segmentation. With-
Related Content

Two Routes to Trust Calibration: Effects of Reliability and Brand Information on Trust in Automation
[www.irma-international.org/article/two-routes-to-trust-calibration/231606/](http://www.irma-international.org/article/two-routes-to-trust-calibration/231606/)

The Impact of Emerging Technology on Leadership Development
[www.irma-international.org/chapter/the-impact-of-emerging-technology-on-leadership-development/226649/](http://www.irma-international.org/chapter/the-impact-of-emerging-technology-on-leadership-development/226649/)

Web 2.0 and Idiosyncrasy of Cultural Heritage: A Perspective from Indonesia
[www.irma-international.org/chapter/web-idiosyncrasy-cultural-heritage/50285/](http://www.irma-international.org/chapter/web-idiosyncrasy-cultural-heritage/50285/)

The Impact of the Model Minority Culture in Higher Education Institutions: The Cause of Asian Americans' Psychological and Mental Health
[www.irma-international.org/chapter/the-impact-of-the-model-minority-culture-in-higher-education-institutions/162530/](http://www.irma-international.org/chapter/the-impact-of-the-model-minority-culture-in-higher-education-institutions/162530/)

The Lessons of Human Resource in The Theory of Constraints