ABSTRACT

Video processing and segmentation are important stages for multimedia data mining, especially with the advance and diversity of video data available. The aim of this chapter is to introduce researchers, especially new ones, to the “video representation, processing, and segmentation techniques”. This includes an easy and smooth introduction, followed by principles of video structure and representation, and then a state-of-the-art of the segmentation techniques focusing on the shot-detection. Performance evaluation and common issues are also discussed before concluding the chapter.

I. INTRODUCTION

With the advances, which are progressing very fast, in the digital video technologies and the wide availability of more efficient computing resources, we seem to be living in an era of explosion in digital video. Video data are now widely available, and being easily generated, in large volumes. This is not only on the professional level. It can be found everywhere, on the internet, especially with the video uploading and sharing sites, with the personal digital cameras and camcorders, and with the camera mobile phones that became almost the norm.

People use the existing easy facilities to generate video data. But at some point, sooner or later, they realize that managing these data can be a bottleneck. This is because the available techniques and tools for accessing, searching, and retrieving video data are not on the same level as for other traditional
data, such as text. The advances in the video access, search, and retrieval techniques have not been progressing with the same pace as the digital video technologies and its generated data volume. This could be attributed, at least partly, to the nature of the video data and its richness, compared to text data. But it can also be attributed to the increase of our demands. In text, we are no longer just satisfied by searching for exact match of sequence of characters or strings, but need to find similar meanings and other higher level matches. We are also looking forward to do the same on video data. But the nature of the video data is different.

Video data is more complex and naturally larger in volume than the traditional text data. They usually combine visual and audio data, as well as textual data. These data need to be appropriately annotated and indexed in an accessible form for search and retrieval techniques to deal with it. This can be achieved based on either textual information, visual and/or audio features, and more importantly on semantic information. The textual-based approach is theoretically the simplest. Video data need to be annotated by textual descriptions, such as keywords or short sentences describing the contents. This converts the search task into the known area of searching in the text data, where the existing relatively advanced tools and techniques can be utilized. The main bottleneck here is the huge time and effort that are needed to accomplish this annotation task, let alone any accuracy issues. The feature-based approach, whether visual and/or audio, depends on annotating the video data by combinations of their extracted low-level features such as intensity, color, texture, shape, motion, and other audio features. This is very useful in doing a query-by-example task. But still not very useful in searching for specific event or more semantic attributes. The semantic-based approach is, in one sense, similar to the text-based approach. Video data need to be annotated, but in this case, with high-level information that represents the semantic meaning of the contents, rather than just describing the contents. The difficulty of this annotation is the high variability of the semantic meaning, of the same video data, among different people, cultures, and ages, to name just a few. It will depend on so many factors, including the purpose of the annotation, the domain and application, cultural and personal views, and could even be subject to the mood and personality of the annotator. Hence, generally automating this task is highly challenging. For specific domains, carefully selected combinations of the visual and/or audio features correlate to useful semantic information. Hence, the efficient extraction of those features is crucial to the high-level analysis and mining of the video data.

In this chapter, we focus on the core techniques that facilitate the high-level analysis and mining of the video data. One of the important initial steps in segmentation and analysis of video data is the shot-boundary detection. This is the first step in decomposing the video sequence to its logical structure and components, in preparation for analysis of each component. It is worth mentioning that the subject is enormous and this chapter is meant to be more of an introduction, especially for new researchers. Also, in this chapter, we only focus on the visual modality of the video. Hence, the audio and textual modalities are not covered.

After this introductory section, section II provides the principles of video data, so that we know the data that we are dealing with and what does it represent. This includes video structure and representation, both for compressed and uncompressed data. The various types of shot transitions are defined in section III, as well as the various approaches of classifying them. Then, in section IV, the key categories of the shot-boundary detection techniques are discussed. First, the various approaches of categorizing the shot-detection techniques are discussed, along with the various factors contributing to that. Then, a selected hierarchical approach is used to represent the most common techniques. This is followed by discussion of the performance evaluation measures and some common issues. Finally the chapter is summarized and concluded in section V.
Related Content

Agile Modeling, Agile Software Development, and Extreme Programming: The State of Research
[www.igi-global.com/article/agile-modeling-agile-software-development/3343?camid=4v1a](www.igi-global.com/article/agile-modeling-agile-software-development/3343?camid=4v1a)

E-R Approach to Distributed Heterogeneous Database Systems for Integrated Manufacturing
[www.igi-global.com/article/approach-distributed-heterogeneous-database-systems/51107?camid=4v1a](www.igi-global.com/article/approach-distributed-heterogeneous-database-systems/51107?camid=4v1a)

3-D Virtual Worlds in Education: Applications, Benefits, Issues, and Opportunities
[www.igi-global.com/chapter/virtual-worlds-education/8053?camid=4v1a](www.igi-global.com/chapter/virtual-worlds-education/8053?camid=4v1a)

The Expert’s Opinion
[www.igi-global.com/article/expert-opinion/51077?camid=4v1a](www.igi-global.com/article/expert-opinion/51077?camid=4v1a)