Chapter IX

Innovative Shot Boundary Detection for Video Indexing

Shu-Ching Chen, Florida International University, USA
Mei-Ling Shyu, University of Miami, USA
Chengcui Zhang, University of Alabama at Birmingham, USA

ABSTRACT

Recently, multimedia information, especially video data, has been made overwhelmingly accessible with the rapid advances in communication and multimedia computing technologies. Video is popular in many applications, which makes the efficient management and retrieval of the growing amount of video information very important. Toward such a demand, an effective video shot boundary detection method is necessary, which is a fundamental operation required in many multimedia applications. In this chapter, an innovative shot boundary detection method using an unsupervised segmentation algorithm and the technique of object tracking based on the segmentation mask maps is presented. A series of experiments on various types of video types are performed, and the experimental results show that our method can obtain object-level information of the video frames as well as accurate shot boundary detection, which are very useful for video content indexing.
INTRODUCTION

Unlike traditional database systems that have text or numerical data, a multimedia database or information system may contain different media such as text, image, audio, and video. Video, in particular, has become more and more popular in many applications such as education and training, video conferencing, video-on-demand (VOD), and news services. The traditional way for the users to search for certain content in a video is to fast-forward or rewind, which are sequential processes, making it difficult for the users to browse a video sequence directly based on their interests. Hence, it becomes important to be able to organize video data and provide the visual content in compact forms in multimedia applications (Zabih, Miller, & Mai, 1995).

In many multimedia applications such as digital libraries and VOD, video shot boundary detection is fundamental and must be performed prior to all other processes (Shahraray, 1995; Zhang & Smoliar, 1994). A video shot is a video sequence that consists of continuous video frames for one action, and shot boundary detection is an operation to divide the video data into physical video shots. Many video shot boundary detection methods have been proposed in the literature. Most of them use low-level global features in the matching process between two consecutive frames for shot boundary detection, for example, using the luminance pixel-wise difference (Zhang, Kankanhalli, & Smoliar, 1993), luminance or color histogram difference (Swanberg, Shu, & Jain, 1993), edge difference (Zabih et al., 1995), and the orientation histogram (Ngo, Pong, & Chin, 2000). However, these low-level features cannot provide satisfactory results for shot boundary detection since luminance or color is sensitive to small changes. For example, Yeo and Liu (1995) proposed a method that uses the luminance histogram difference of DC images, which is very sensitive to luminance changes. There are also approaches focusing on the compressed video data domain. For example, Lee, Kim, and Choi (2000) proposed a fast scene/shot change detection method, and Hwang and Jeong (1998) proposed the directional information retrieving method by using the discrete cosine transform (DCT) coefficients in MPEG video data.

In addition, dynamic and adaptive threshold determination is also applied to enhance the accuracy and robustness of the existing techniques in shot cuts detection (Alattar, 1997; Gunes, Ferman, & Tekalp, 1998; Truong, Dorai, & Venkatesh, 2000). In Gunes et al. (1998), the unsupervised clustering algorithm proposed a generic technique that does not need threshold setting and allows multiple features to be used simultaneously; while an adaptive threshold determination method that reduces the artifacts created by noise and motion in shot change detection was proposed by Truong et al. (2000).

In this chapter, we present an innovative shot boundary detection method using an unsupervised image-segmentation algorithm and the object-tracking technique on the uncompressed video data. In our method, the image-segmentation algorithm extracts the segmentation mask map of each video frame automatically, which can be deemed as the clustering feature map of each frame and where the pixels in each frame have been grouped into different classes (e.g., two classes). Then the difference between the segmentation mask maps of two frames is checked. Moreover, due to camera panning and tilting, we propose an object-tracking method based on the segmentation results to enhance the matching. The cost for object tracking is almost trivial since the segmentation results are already available. In addition, the bounding boxes and the positions of
Retrieval of Web Pages on Real-World Events related to Physical Objects
*International Journal of Information Retrieval Research* (pp. 65-80).
[www.igi-global.com/article/retrieval-web-pages-real-world/72706?camid=4v1a](www.igi-global.com/article/retrieval-web-pages-real-world/72706?camid=4v1a)