Chapter 11
Video Segmentation and Structuring for Indexing Applications

Ruxandra Tapu
TELECOM SudParis, France

Titus Zaharia
TELECOM SudParis, France

ABSTRACT

This paper introduces a complete framework for temporal video segmentation. First, a computationally efficient shot extraction method is introduced, which adopts the normalized graph partition approach, enriched with a non-linear, multiresolution filtering of the similarity vectors involved. The shot boundary detection technique proposed yields high precision (90%) and recall (95%) rates, for all types of transitions, both abrupt and gradual. Next, for each detected shot, the authors construct a static storyboard by introducing a leap keyframe extraction method. The video abstraction algorithm is 23% faster than existing techniques for similar performances. Finally, the authors propose a shot grouping strategy that iteratively clusters visually similar shots under a set of temporal constraints. Two different types of visual features are exploited: HSV color histograms and interest points. In both cases, the precision and recall rates present average performances of 86%.

INTRODUCTION

Recent advances in the field of image/video acquisition and storing devices have determined a spectacular increase of the amount of audio-visual content transmitted, exchanged and shared over the Internet. In the past years, the only method of searching information in multimedia databases was based on textual annotation, which consists of associating a set of keywords to each individual item. Such a procedure requires a huge amount of human interaction and is intractable in the case of large multimedia databases. Today, existing video repositories (e.g., Youtube, Google Videos, DailyMotion, etc.) include millions of items. Thus, attempting to manually annotate...
Video Segmentation and Structuring for Indexing Applications

such huge databases is a daunting job, not only in terms of money and time, but also with respect to the quality of annotation.

When specifically considering the issue of video indexing and retrieval applications, because of the large amount of information typically included in a video document, a first phase that needs to be performed is to structure the video into its constitutive elements: chapters, scenes, shots and keyframes. This paper specifically tackles the issue of video structuring and proposes a complete and automatic segmentation methodology.

Figure 1 presents the proposed analysis framework. The main contributions proposed in this paper concern: an enhanced shot boundary detection method, a fast static storyboard technique and a new scene/chapter detection approach.

The rest of this paper is organized as follows. After a brief recall of some basic theoretical aspects regarding the graph partition model exploited, we introduce the proposed shot detection algorithm. Then, we describe the keyframe selection procedure. The following section introduces a novel scene/chapter extraction algorithm based on temporal distances and merging strategies. The experimental results obtained are then presented and discussed in details. Finally, we conclude the paper and open some perspectives of future work.

SHOT BOUNDARY DETECTION

Related Work

The first methods introduced in the literature were based on pixels color variation between successive frames (Zhang et al., 1993; Lienhart et al., 1997). Such algorithms offer the advantage of simplicity but present serious limitations. Thus, in the presence of large moving objects or in the case of camera motion, a significant number of pixels change their intensity values, leading to false alarms. In addition, such methods are highly sensitive to noise that may be introduced during the acquisition process.

Among the simplest, most effective and common used methods, the color histogram comparison and its numerous variations assume that frames from the same shot have similar histograms (Yuan et al., 2007; Gargi et al., 2000). Histogram-based methods are more robust to noise and motion than pixel-based approaches due to the spatial invariance properties. However, they also present some strong limitations. First of all, let us mention the sensitivity to abrupt changes in light intensity: two images taken in the same place but with different lightening conditions will be described by distinct histograms. Furthermore, a color histogram does not take into account the spatial information of an image, so two identical histograms could actually correspond to two visual completely different images, but with similar colors and appearance probability (Matsumoto et al., 2006).

Let us also cite the methods based on edges/contours (Zabih et al., 1995). Such methods are useful in removing false alarms caused by abrupt illumination change, since they are less sensitive to light intensity variation than color histogram