Chapter VIII

Video Shot Boundary Detection

Waleed E. Farag, Zagazig University, Egypt
Hussein Abdel-Wahab, Old Dominion University, USA

ABSTRACT

The increasing use of multimedia streams nowadays necessitates the development of efficient and effective methodologies for manipulating databases storing this information. Moreover, in its first stage, content-based access to video data requires parsing of each video stream into its building blocks. The video stream consists of a number of shots, each one a sequence of frames pictured using a single camera. Switching from one camera to another indicates the transition from a shot to the next one. Therefore, the detection of these transitions, known as scene change or shot boundary detection, is the first step in any video-analysis system. A number of proposed techniques for solving the problem of shot boundary detection exist, but the major criticisms to them are their inefficiency and lack of reliability. The reliability of the scene change detection stage is a very significant requirement because it is the first stage in any video retrieval system; thus, its performance has a direct impact on the performance of all other stages. On the other hand, efficiency is also crucial due to the voluminous amounts of information found in video streams. This chapter proposes a new robust and efficient paradigm capable of detecting scene changes on compressed MPEG video data directly. This paradigm constitutes the first part of a Video Content-based Retrieval (VCR) system that has been designed at Old Dominion University. At first, an abstract representation of the compressed video stream, known as the DC
sequence, is extracted, then it is used as input to a Neural Network Module that performs the shot boundary-detection task. We have studied experimentally the performance of the proposed paradigm and have achieved higher shot boundary detection and lower false alarms rates than other techniques. Moreover, the efficiency of the system outperforms other approaches by several times. In short, the experimental results show the superior efficiency and robustness of the proposed system in detecting shot boundaries and flashlights — sudden lighting variation due to camera flash occurrences — within video shots.

INTRODUCTION

The recent explosive growth of digital video applications entails the generation of vast amount of video data; however, the technologies for organizing and searching video databases are still in their infancy. The first step in indexing video databases (to facilitate efficient access) is to analyze the stored video streams. Video analysis can be classified into two stages (Rui, Huang, & Mehrotra, 1998): shot boundary detection and key frames extraction. The purpose of the first stage is to partition a video stream into a set of meaningful and manageable segments, whereas the second stage aims to abstract each shot using one or more representative frames. We will address the problem of shot boundary detection in this chapter while the problem of selecting key frames from segmented shots will be dealt with in another chapter.

In general, successive frames in motion pictures bear great similarity among themselves, but this generalization is not true at the boundaries of shots. A frame at a boundary point of a shot differs in background and content from its successive frame that belongs to the next shot (Figure 1). In a nutshell, two frames at a boundary point will differ significantly as a result of switching from one camera to another, and this is the basic principle upon which most automatic algorithms for detecting scene changes depend.

Due to the huge amount of data contained in video streams, almost all of them are transmitted and stored in compressed format. While there are large numbers of algorithms for compressing digital video, the MPEG format (ISO/IEC, 1999; LeGall, 1991; Mitchell, Pennebaker, Fogg, & LeGall, 1997) is the most famous one and the current international standard. In MPEG, spatial compression is achieved through the use of a Discrete Cosine Transform (DCT)-based algorithm similar to the one used in the JPEG standard (Pennebaker & Mitchell, 1993; Wallace, 1991). In this algorithm, each frame is divided into a number of blocks (8X8 pixel), then the DCT transformation is applied to these blocks. The produced coefficients are then quantized and entropy-encoded, a technique that achieves the actual compression of the data. On the other side, temporal compression is accomplished using a motion compensation technique that depends on the similarity between successive frames on video streams. Basically, this technique codes the first picture of a video stream (I frame) without reference to neighboring frames, while successive pictures (P or B frames) are generally coded as differences to that reference frame(s). Considering the large amount of processing power required in the manipulation of raw digital video, it becomes a real advantage to work directly upon compressed data and avoid the need to decompress video streams before manipulating them.
Related Content

Concept of Association Rule of Data Mining Assists Mitigating the Increasing Obesity

An Efficient and Flexible Approach of Keyword Search in XML
[www.irma-international.org/chapter/efficient-flexible-approach-keyword-search/69125/](http://www.irma-international.org/chapter/efficient-flexible-approach-keyword-search/69125/)

Ranking Algorithm for Semantic Document Annotations
[www.irma-international.org/article/ranking-algorithm-semantic-document-annotations/72703/](http://www.irma-international.org/article/ranking-algorithm-semantic-document-annotations/72703/)

Service-Driven Computing: Challenges and Trends
[www.irma-international.org/chapter/service-driven-computing/198648/](http://www.irma-international.org/chapter/service-driven-computing/198648/)

A Cross-Cultural Measure of Servant Leadership Behaviors
[www.irma-international.org/chapter/cross-cultural-measure-servant-leadership/69739/](http://www.irma-international.org/chapter/cross-cultural-measure-servant-leadership/69739/)