Chapter 9
A Comparative Study of Unsupervised Video Shot Boundary Detection Techniques Using Probabilistic Fuzzy Entropy Measures

Biswanath Chakraborty
RCC Institute of Information Technology, India

Siddhartha Bhattacharyya
RCC Institute of Information Technology, India

Susanta Chakraborty
Bengal Engineering and Science University, India

ABSTRACT

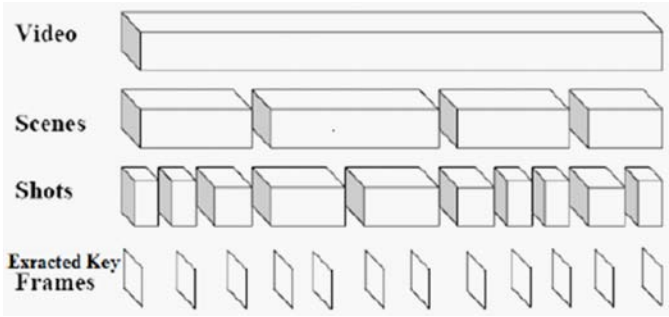
The performance of video shot boundary detection technique in unsupervised video sequence can be improved by the use of different probabilistic fuzzy entropies. In this chapter, the authors present a new technique for identifying as to whether there are any appreciable changes from one video context to another in the available sequence of image frames extracted from a mixture of a numbers of video files. They then compared their technique with an existing technique and found improved performance of the video shot boundary detection techniques using probabilistic fuzzy entropies.

INTRODUCTION

A video consists of a several images/frames which are played consecutively at a constant speed of around 20 to 30 frames per second for smooth visualization (Thakar. *et al*). Video shot boundary detection means to segment a video by detecting boundaries between two camera shots (frames) which are the first and the most important step for content-based video retrieval. A video can be divided in the hierarchy (Figure 1) as follows: video → scenes → shots → frames →pixels (Das. *et. al* 2008). So, a shot can be defined as consecutive frames recorded from a single camera operation, representing a continuous action in

DOI: 10.4018/978-1-4666-2518-1.ch009

Figure 1. Overview of shot boundary detection



time and space and typically relevant shots are grouped into a higher level unit called a scene. Shot boundary detection means to find the locations of shot transitions, which are mainly of two types, viz., cut and gradual transitions. Gradual shot transition occurs over a few video frames. Fade in/out and dissolves are the two most common gradual shot transitions. A fade is a gradual transition between a scene and a constant image which is called fade out and that between a constant image and a scene which is called fade in. A dissolve is a gradual transition from one scene to another (where two consecutive shots temporally superimpose).

Video content classification is an important task in the computer vision community as far as intelligent analysis of video content is concerned. Typical applications include video content mining, video surveillance and defense applications to name a few. A score of literatures exists in this regard providing a detailed comparative study of the different techniques in vogue (Boreczky *et al*. 1996). Video content classification primarily entails the detection of the cuts in the video sequences through classical as well as non-classical techniques. The non-classical techniques resort to computational intelligence perspectives in ascertaining the video shot boundaries by using the inherent information distribution of the video content. Hence, larger video frames/sequences obviously require a lot of computational overhead in this shot detection task. Since, video data are essentially continuous in nature, dynamic estimation of the data distribution is a prerequisite in the commonly used shot boundary detection and video analysis techniques. The commonly used estimation techniques include:

- *Using the frequency domain* Porter S.V. *et al.* (S.V. *et al.*) have shown by using the average interframe correlation coefficients to determine whether an abrupt shot change has occurred, and
- *Using edge variance* Yoo *et al.*(Yoo *et al.*, 2006) describe a new method to detect gradual transitions here edge variance as feature and compared with local blocks in the average edge frames. In addition, the variance curves are smoothed by an opening operation. This algorithm compares each frame with some adjacent frames to get a systematic view of relationship in the video sequence.
- *Using the spatial and the frequency component information* of the moving objects (Jacobson *et al.* 1987, Heeger 1988). In (Jacobson *et al.* 1987), Jacobson and Wechsler have shown the usefulness of spatio-temporal frequency measure for determining optical flow field in a video environment and Heeger (Heeger 1988) proposed filters based on spatio-temporal features for the same purpose viz. determination of optical flow field, and
- *Using non-linear approach* Hua Kein A., *et al.* (Hua *et al.*, 2000) proposed a tech-