

# An Efficient Method for Detection of Key Objects in Video Shots with Camera Motions

JungHwan Oh   JeongKyu Lee   Sae Hwang \*

## Abstract

The most fundamental task in video processing is to partition long video sequences into a number of shots and find a *key frame* of each shot for indexing and browsing. In this paper, we extract *key objects* instead of key frames. In order to facilitate this process, we extend our previous technique for shot boundary detection (SBD) to select two candidate frames out of each shot, then propose a new framework to segment key objects from those selected frames using *color quantization* and *background adjustment*. We make our scheme cost-effective and automatic by avoiding expensive computations, and removing manual processing. Experimental results show that the proposed scheme is very promising.

**Keywords:** *Video object segmentation, shot boundary detection, color quantization, MPEG-4/MPEG-7.*

## 1 Introduction

With the rapid advances in networking and coding technologies, video has become an essential part of many important applications such as digital libraries, distance learning, public information systems, electronic commerce, and entertainment. However, due to the enormous size of video files and their semantically rich contents, organizing and managing video as data are much more complex than manipulating text.

Video obtained from various types of sources is called a *video clip* with varying lengths of time lasting from a few seconds to several hours. For most video applications a video clip is not a convenient unit for data entry since an entire video stream is too coarse as a level of abstraction. There is a need for a new basic unit to handle video data. For the purposes of this scheme, it has been agreed that *shot*, which is defined as a collection of frames recorded from a single camera operation, is the winner among the several candidates since shot boundaries can be decided objectively and mechanically.

Several research projects [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] have been performed to find these shot boundaries automatically, in other words, to parse a long video into a number of shots which are the basic units for video processing. In our previous works [11, 12, 13, 14, 15], we proposed the shot boundary detection (SBD) technique, which processes substantially less data since it uses only backgrounds instead of whole area of images. Also it is much less sensitive to the threshold values, and very effective in reducing incorrect detection. After a video is decomposed into shots, usually, the next step is to extract a *key frame* from each shot for content based summarizing and understanding. In this paper, we extract *key objects*

---

\*Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019-0015 U. S. A. e-mail:{oh, jelee, hwang}@cse.uta.edu