

Structuring and indexing of digital video by color histogram shot boundary detection

THIAGO TEIXEIRA SANTOS^{1,2}, CARLOS HITOSHI MORIMOTO¹

¹IME–Instituto de Matemática e Estatística - Rua do Matão, 1010, 05508-090 São Paulo, SP, Brasil
{thsant, hitoshi}@ime.usp.br

²Faculdades Senac - Rua Tito, 54, 05051-000 São Paulo, SP, Brasil

Abstract. Video contains huge amounts of information which needs to be structured and indexed to make tasks like search and navigation possible. This paper describes ongoing research for the development of a prototype system to segment digital video in *shots* and index them by *key-frames*. Preliminary results of videos structured as a sequence of key-frames is available online.

1 Introduction

Digital video provides us with powerful ways for video representation, storage, distribution and annotation, such as those seen in MPEG-4 and MPEG-7. But to facilitate the access to video information it is necessary to create new tools to automatically structure and retrieve video content.

The simplest unit of a video structure, beyond the frame level, is the *shot*. A shot consists of a contiguous sequence of frames that represents continuous action in time and space, acquired in a single camera recording operation. During editing, shots are assembled to make the full video. The transition between shots can be abrupt, called a cut, or smooth, where frames from different shots are overlapped. Some examples of smooth transitions are fades, dissolves and wipes.

This work presents a prototype for indexing and structuring video according to shots. Shots have higher granularity than video frames, resulting in a more efficient way for browsing and searching video data [1].

2 Structuring and Indexing

Shot Boundary Detection (SBD) is the problem of computing the frames that lie on the boundaries (start and finish) of each shot. For example, SBD could be used to segment an analog video, after digitization, to recover the original shots used for editing. Several methods have been proposed in the literature to solve SBD [2]. The technique used in our prototype is based on the *twin-comparison* method introduced by Zhang et al. [2], which uses the difference of histograms between frames. The difference signal is filtered by median subtraction [3], before the computation of the boundaries.

Once the boundaries are computed, a single key-frame is used to represent each shot. The key-frame is selected using a heuristic algorithm that computes the frame with higher temporal support. The sequence of key-frames is an efficient way of representing the video information, similar to a story board or comic book, as seen in Figure 1. This structure also creates a natural index that can be used

to directly access one or a collection of similar shots by just clicking at the key-frame. We are extending the prototype with more functionalities, and experimenting new SBD techniques that are more robust to smooth transitions. More preliminary results are available in our site [4].

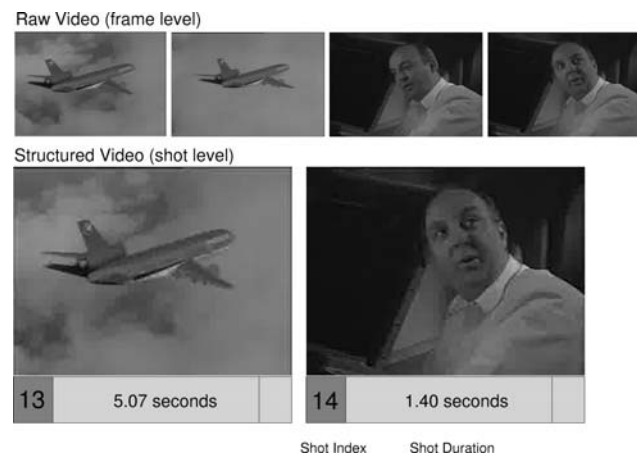


Figure 1: Automatic shot level structuring and indexing.

References

- [1] G. Davenport, T. A. Smith and N. Pincever, *Cinematic Primitives for Multimedia*, IEEE Computer Graphics and Applications, Volume 11 (1991).
- [2] I. Koprinska e S. Carrato, *Temporal video segmentation: A survey*, Signal Processing: Image Communication, Volume 16 (2001).
- [3] A. M. Ferman and A. M. Tekalp, *Efficient Filtering and Clustering Methods for Temporal Video Segmentation and Visual Summarization*, Journal of Visual Communication and Image Representation (1998).
- [4] <http://www.ime.usp.br/~thsant/vweb.html>