

# A Hierarchical Organization of Home Video

**Yu-Jin Zhang**

*Department of Electric Engineering, Tsinghua University, China*

## INTRODUCTION

Video analysis is an important branch of content-based video retrieval (CBVR). Home video (also called personal video) is an umbrella term used to indicate video captures by non-professional persons (Zhang, 2003). With the progress of electronic equipments and computer technology for taking motion pictures and processing huge data, many people now own and use their personal camcorders to make home videos that capture their everyday experiences and document their lives. Home video has no time limits and no restriction in content (Lienhart, 2000), so these videos easily add up to many hours of material. The organization and edition of home videos present technical challenges due to the lack of efficient tools (Arora et al., 2013). Though a number of prototype systems for content-based video analysis and retrieval have been constructed, for example, as shown in (Wactlar et al., 1996; Chang et al., 1998), the development of tools and systems specialized for addressing home video, that is for extracting, representing, organizing, browsing, querying and retrieving video, is just on a preliminary stage (Huang et al., 2005; Wu & Obrador, 2005).

Several tasks are needed to confront the feasible organization of home video. Home video has certain particular characteristics. The organization of home video should be based on the understanding of video structures and contents, and by taking advantages of the structures and contents. Home video are completed and stored straight in compressed domain. In order to save both time and space, techniques that manipulate home videos directly in compressed domain should be considered (Wang et al., 2003). Some typical techniques working on compressed domain could be found in (Taskiran et al., 2004). Home video are made by shot after shot without storyline, these shots may or may not have immediate relationship. To group shots, the visual features should be extracted from every shot (Gatica-Perez et al., 2003).

Facing the above-mentioned tasks and difficulties, a novel technique for such a home video organization is described in this article. It is based on the analysis of characteristics of home video, on the detection of motion attention regions in compressed domain, on the time weighting model based on camera motion for attention guiding, on a novel two-layer shot clustering approach and organization strategy, and on an efficient shot clustering algorithm. These enabling techniques will be explained in the following. Experiments made on two home videos from MPEG-7 data set with encouraging results are also presented. Finally, further research directions and concluding remarks are provided.

## BACKGROUND

Compared to other types of video programs, home video has some particularities according to the persons in shoot and objects to be screened (Lienhart et al., 1997). The study on home video analysis may benefit from its unique characteristics.

In general, a typical home video has certain structure characteristics: it contains a set of scenes, each composed of ordered and temporally adjacent shots that can be organized in clusters conveying semantic meaning. The fact is that home video recording imposes temporal continuity. Unlike other video programs, home video just records the life but not composes story, so every shot (clip of video captured in one place without interruption) may have the equal importance. In addition, filming home video with a temporal back-and-forth structure is rare. For example, on a vacation trip, people do not usually visit the same site twice. In other words, the content tends to be localized in time. Consequently, discovering the scene structure above shot level plays a key role in home video analysis. Video content organization based on shot clustering provides an efficient way of semantic video accessing and fast video editing.

DOI: 10.4018/978-1-4666-5888-2.ch210

Home video is not prepared for very large audience (like broadcasting TV), but for relatives, guests and friends. To analyze home video, the purpose and filming tact should be considered. For example, the subjective feeling transferred by video information can be decomposed into two parts: one from motion region that attract the attention of viewers, another from the general impression of environment. In the same time, different types of camera motions should also be considered. Different camera motions may signify different changes of attentions. For example, zoom in makes the attention of viewers more on motion regions, while pan and tilt make the notice of viewers more on environments.

Home video may be made by different persons in different circumstances. How to reflect the viewers' visual perception during the filming needs to be considered. In fact, different viewer's attention will never stay equal all through the watching of home video. The importance of certain frames or clips should be weighted reasonably by their relative importance according to subjective perception. Sometimes, users are required to assign these weights to express their real preference (Chang et al., 1998; Babu & Ramakrishnan, 2002). Other approaches include temporally making weights in key frame selection procedure by assuming that the greater number of skipped frames, the more weight is assigned to the current frame (Tan et al., 1999), or building up a generic user attention model by integrating a set of audio-visual attention model features extracted from video sequence (Ho et al., 2003).

## ATTENTION REGION DETECTION

Structuring and organizing video needs the detection of motion information, here represented as attention regions. This is not equal to the detection of video objects. The detection of video objects requires accurately determination of the boundary of objects and quickly following of the change of objects. For example, an object-based approach first imposes spatial-temporal segmentation to get individual regions (Achanta et al., 2002). The video structuring stresses more on the subjective feeling of human beings in viewing video. In this regard, the influence of region detection on subjective feeling is more important than just accurate segmentation (Zhai & Shah, 2005).

When there was no specific clear-shaped object in the background, then most object detection methods would be failed. In order to circumvent the problem of actual object segmentation, a different concept, "attention region," has been proposed and be employed (Jiang & Zhang, 2005). An attention region does not necessarily correspond to a real object, but denotes the region with irregular movement compared to camera motion. Based on the assumption that different movement from the global motion attracts more attention (supported by the common sense that irregular motion tends to be easily caught by human eyes in a static or regular moving background), these regions are regarded as somewhat important areas in contrast to the background image. The detection of attention region requires less precision in general, since more emphasis has been placed on the degree of human attention than on the accurate object outline. This task could be simply completed by detecting the outliers in frame. In addition, the tracking process for attention region is easier, too.

The first step of attention region detection is to segment a "dominant region" from a single frame, which is illustrated in Figure 1. Figure 1(a) is a typical home video frame. This frame can be decomposed into two parts: the running boy on grassplot, as shown in Figure 1(b); and the background grasses and trees, as shown in Figure 1(c). The latter represents the environment and the former corresponds to the attention region.

The detection of attention regions does not require at very high precision. Therefore, the detection of attention regions can be performed directly in MPEG compressed domain. Two types of information in compressed domain can be used (Jiang & Zhang, 2005):

1. **DCT Coefficients of Macro Block:** Among DCT coefficients, the DC coefficient is easy to be obtained. It is the directly obtainable component of image block, and its value equals to eight time of the average value in block. It roughly reflects the brightness and color information.
2. **Motion Vectors of Macro Block:** Motion vectors correspond to sparse and coarse motion field. They reflect approximately the general motion information in the block.

With the motion vectors, a simple but effective four-parameter global motion model can be used to

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/a-hierarchical-organization-of-home-video/112627](http://www.igi-global.com/chapter/a-hierarchical-organization-of-home-video/112627)

## Related Content

---

### Radio Frequency Identification Systems Within a Lean Supply Chain in a Global Environment

Alan D. Smith, Terry Stringer Damron, Susan Cockrelland Amye M. Melton (2018). *Encyclopedia of Information Science and Technology, Fourth Edition* (pp. 5516-5526).

[www.irma-international.org/chapter/radio-frequency-identification-systems-within-a-lean-supply-chain-in-a-global-environment/184253](http://www.irma-international.org/chapter/radio-frequency-identification-systems-within-a-lean-supply-chain-in-a-global-environment/184253)

### Maturity for Sustainability in IT: Introducing the MITS

Martijn Smeitinkand Marco Spruit (2013). *International Journal of Information Technologies and Systems Approach* (pp. 39-56).

[www.irma-international.org/article/maturity-sustainability-introducing-mits/75786](http://www.irma-international.org/article/maturity-sustainability-introducing-mits/75786)

### Capacity for Engineering Systems Thinking (CEST): Literature Review, Principles for Assessing and the Reliability and Validity of an Assessing Tool

Moti Frank (2009). *International Journal of Information Technologies and Systems Approach* (pp. 1-14).

[www.irma-international.org/article/capacity-engineering-systems-thinking-cest/2543](http://www.irma-international.org/article/capacity-engineering-systems-thinking-cest/2543)

### Manipulator Control Based on Adaptive RBF Network Approximation

Xindi Yuan, Mengshan Liand Qiusheng Li (2023). *International Journal of Information Technologies and Systems Approach* (pp. 1-16).

[www.irma-international.org/article/manipulator-control-based-on-adaptive-rbf-network-approximation/326751](http://www.irma-international.org/article/manipulator-control-based-on-adaptive-rbf-network-approximation/326751)

### Distance Teaching and Learning Platforms

Linda D. Grooms (2018). *Encyclopedia of Information Science and Technology, Fourth Edition* (pp. 2455-2465).

[www.irma-international.org/chapter/distance-teaching-and-learning-platforms/183958](http://www.irma-international.org/chapter/distance-teaching-and-learning-platforms/183958)