Organization of Home Video

Yu-Jin Zhang
Tsinghua University, Beijing, China

INTRODUCTION

With the progress of electronic equipments and computer technology for taking motion pictures and processing huge data, an increasing number of people now own and use camcorders to make home videos that capture their 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. However, the organization and edition of the large amount of information contained in home videos present technical challenges due to the lack of efficient tools. Though a number of prototype systems for content-based video analysis and retrieval have been constructed, for example, as shown in (Wactlar, 1996; Chang, 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, 2005; Wu, 2005).

Several tasks are needed to confront to make the organization of home video possible and feasible. Home video has certain particular characteristics. The organization of home video should be based on the understanding of video structures, and by taking advantages of this structure. 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, 2003). Some typical techniques working on compressed domain could be found in (Taskiran, 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, 2003)

Facing these tasks and difficulties, a novel technique 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 based on camera motion, and on a novel two-layer shot clustering approach and organization strategy. Experiments made on two home videos from MPEG-7 data set provide encouraging results.

BACKGROUND

Video analysis is an important branch of content-based video retrieval (CBVR). Compared to other types of video programs, home video has some particularities according to the persons in shoot and objects to be screened [Lienhart 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.

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, 1998; Babu, 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, 1999), or building up a generic user attention model by integrating a set of audio-visual attention model features extracted from video sequence (Ho, 2003).

MAIN FOCUS OF THE CHAPTER

Detection of Attention Regions

Structuring video needs the detection of motion 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, 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, 2005).

It is know that most object detection methods would fail if there were no specific clear-shaped object in the background. In order to circumvent the problem of actual object segmentation, a different concept: “attention region” could be employed. 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, 2005):

DCT Coefficients of Macro Block

Among DCT coefficients, DC coefficient is easy to get. 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.

Motion Vectors of Macro Block

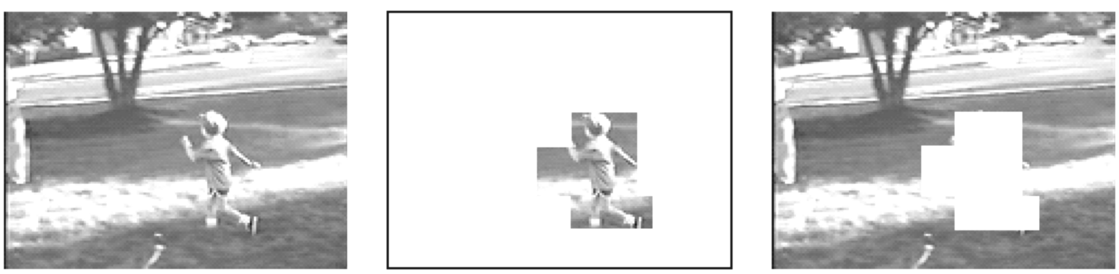
Motion vectors correspond to sparse and coarse motion field. They reflect approximately the general motion information in the block.

With the motion vectors, a simple but effective four-parameter global motion model can be used to estimate simplified camera motions (zoom, rotate, pan and tilt):

$$\begin{cases} u = h_1 x + h_2 y + h_3 \\ v = h_2 x + h_1 y + h_4 \end{cases} \tag{1}$$

A common least-square fitting algorithm is imposed to optimize the model parameters h_1, h_2, h_3, and h_4. This algorithm recursively examines the error produced by the current estimate of the camera parameters and generates an outlier mask, consisting of macro-blocks with motion vectors not following the camera motion model. Then, the camera parameters are re-computed and new outlier mask are formed. This process iterates until the parameters are stabilized. As it operates directly on MPEG video, motion

Figure 1. Illustration of attention region and environment



(a) (b) (c)

Related Content

Energy Management in Wireless Networked Embedded Systems
G. Manimaran (2009). *Encyclopedia of Information Science and Technology, Second Edition (pp. 1381-1386).*
www.irma-international.org/chapter/energy-management-wireless-networked-embedded/13756

Artificial Haemostasis System for Modern Information Retrieval with 3D Result-Mining
Hadj Ahmed Bouarara, Reda Mohamed Hamouand Abdelmalek Amine (2016). *Journal of Information Technology Research (pp. 17-46).*
www.irma-international.org/article/artificial-haemostasis-system-for-modern-information-retrieval-with-3d-result-mining/160155

Case Study Part 2: Apples, Oranges, and Pears
(2018). *Measuring the Validity of Usage Reports Provided by E-Book Vendors: Emerging Research and Opportunities (pp. 30-54).*
www.irma-international.org/chapter/case-study-part-2/190052

Agent-Based Negotiation in E-Marketing
V.K. Murthyand E.V. Krishnamurthy (2009). *Encyclopedia of Information Science and Technology, Second Edition (pp. 88-92).*
www.irma-international.org/chapter/agent-based-negotiation-marketing/13554

TAM Model Evidence for Online Social Commerce Purchase Intention
Zhang Ying, Zeng Jianqiu, Umair Akramand Hassan Rasool (2021). *Information Resources Management Journal (pp. 86-108).*
www.irma-international.org/article/tam-model-evidence-for-online-social-commerce-purchase-intention/270887