Signal Processing: *Image Communication* 28 (2013) 241–253

Contents lists available at SciVerse ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Video abstraction based on the visual attention model and online clustering

Qing-Ge Ji [a], Zhi-Dang Fang [a], Zhen-Hua Xie [a], Zhe-Ming Lu [b,*]

[a] *School of Information Science and Technology, Sun Yat-sen University, Guangzhou 510006, PR China*
[b] *School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, PR China*

ARTICLE INFO

Article history:
Received 31 March 2012
Accepted 22 November 2012
Available online 1 December 2012

Keywords:
Video abstraction
Saliency map
Key frame
Region of interest
Online clustering

ABSTRACT

With the fast evolution of digital video, research and development of new technologies are greatly needed to lower the cost of video archiving, cataloging and indexing, as well as improve the efficiency and accessibility of stored video sequences. A number of methods to respectively meet these requirements have been researched and proposed. As one of the most important research topics, video abstraction helps to enable us to quickly browse a large video database and to achieve efficient content access and representation. In this paper, a video abstraction algorithm based on the visual attention model and online clustering is proposed. First, shot boundaries are detected and key frames in each shot are extracted so that consecutive key frames in a shot have the same distance. Second, the spatial saliency map indicating the saliency value of each region of the image is generated from each key frame and regions of interest (ROI) is extracted according to the saliency map. Third, key frames, as well as their corresponding saliency map, are passed to a specific filter, and several thresholds are used so that the key frames containing less information are discarded. Finally, key frames are clustered using an online clustering method based on the features in ROIs. Experimental results demonstrate the performance and effectiveness of the proposed video abstraction algorithm.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Since video has become an indispensable part in multimedia exchange, what we can access now are not only pure text documents but also multimedia data that combines all types of media such as text, images, audio and video. However, this leads us to the issue of a large amount of data, which increases the cost of storage and transmission. What we face is not a lack of video resources but the requirements to find what we are interested in from a great amount of data fast and precisely enough and to utilize and manage the video data as well as we can. Therefore, how to effectively and rapidly deal with the large quantity of video data in the applications of storage, management, classification, indexing, retrieval and browsing, is an urgent problem that has to be solved. Among all the components for video processing, video abstraction is an essential one for video retrieval and recognition. Since traditional retrieval methods based on keywords and description texts seem not capable of satisfying the new requirements any longer, content-based video retrieval (CBVR) [1,2] techniques have been proposed in the past fifteen years to search the required video data in the whole video database based on the video content. CBVR has been successfully applied to video retrieval, video-on-demand, video meeting, video surveillance, multimedia data processing and family entertainment [3–5].

Video abstraction is the key part in the CBVR framework. There are two fundamentally different types of

* Correspondence to: Room 105, Teaching Building 5, Yuquan Campus, Zhejiang University, Hangzhou 310027, P. R. China.
Tel./fax: +86 571 879 53349.
E-mail addresses: zheminglu@zju.edu.cn, nampl@zju.edu.cn (Z.-M. Lu).

0923-5965/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.image.2012.11.008

video abstraction schemes: video summary and video skimming [6]. The former is known as still-image abstraction or static storyboard, which is a small collection of salient images extracted or generated from the underlying video source. The latter is known as moving-image abstraction or moving storyboard, which consists of a collection of image sequences, as well as the corresponding audio abstraction extracted from the original sequence and is thus itself a video clip but of considerably shorter length. Both of them are required to be consistent with human perception and compact, that is, the abstraction results should be concentrated and able to represent the semantic content of the original video without losing important information. Since traditional retrieval or abstraction methods based on pixels are weak at semantic expression and are inconvenient for users to understand, visual content representations with a higher semantic level are required. Until now a number of research works have been carried out and various algorithms and methods are presented for video abstraction. Quite a few works put emphasis upon static summary, namely key frame representation, because the framework is easier to construct, the extracted key frames offer users an intuitive description, and users are able to browse anywhere that they are interested in based on key frames in the video sequence. A comprehensive overview and classification of these works is presented in [7], and some typical schemes can be introduced as follows. Zhuang et al. [8] proposed a key frame extraction method based on clustering. In their scheme, for each frame, a 16×8 2D HS color histogram based on the HSV color space is calculated. Frames are then clustered according to their histogram similarity. Finally, key frames are selected from each cluster. Recently, Huang [9] has proposed a mutual information based approach. In his approach, the joint probability density and entropy value of two consecutive frames are calculated based on the RGB space. After a curve of mutual information for all consecutive frames is built, frames are clustered according to the mutual information, and thus key frames can be selected. Fayk et al. [10] utilized the particle swarm optimization technique for key frame selection. However, since their method divides the video into equal segments, the segment size has an obvious influence on the results. Kim and Hwang [11] presented an object-based video abstraction scheme, where video object segmentation is used in the framework of video surveillance system to extract the video object plane and successive comparisons are carried out to determine the key VOPs as key frames. Doulamis et al. [12] came out with a new approach for efficient visual content representation. In their scheme, after extracting features from the video sequences using a color/motion segmentation algorithm, a multidimensional fuzzy histogram is constructed. And this method is proved to have a good performance in both video abstraction and content-based retrieval. Li et al. [13] utilized the min-max distortion optimization method to extract the key frames from a video. They addressed the summarization problem by finding a predetermined number of frames that minimize the temporal distortion and adopted dynamic programming techniques to optimize the process. Doulamis et al. [14] regarded the visual content as a feature curve in a high-dimensional feature space and extracted key frames by appropriately selecting representative points in the feature curve. Fu et al. [15] provided a solution to the summarization of multi-view videos by combining features of different shots and using a hypergraph to model the correlations between these shots. Ma et al. [16] introduced a generic attention model and applied it to video summarization. In their scheme, different features, including colors, sound, camera motions and object motions, were discussed for users' attention. Video decompositions can also be used for video abstraction and navigation. Doulamis et al. [17] proposed a content-based video organization scheme. In their work, video sequences were analyzed hierarchically at different content resolution levels to form a tree structure, where the tree nodes indicate the temporal video segments that the sequence content is partitioned at a given resolution. To organize and access video content effectively and fast, several research works on video browsing have been proposed. Zhu et al. [18] proposed a video processing technique to organize the content hierarchy of the video. Zhu and Zhou [19] presented a system for video browsing and retrieval based on multimedia integration. Schoeffmann and Boeszoermenyi [20] described an interactive video browsing scheme using navigation summaries. It allows our random access to a video and provides abstract visualizations of the content at a user-defined level of detail and, thus, quickly communicates content characteristics to the user.

Among various algorithms for video abstraction, such as those introduced above, methods based on feature extraction and clustering are quite common ones, as frames can be organized intuitively and it is convenient to perform information retrieval. However, there are shortcomings on the amount of calculation and the redundancy of information. In most approaches, data in the whole frame is taken into account for feature calculation or similarity comparison. And all the regions in the frame are treated equally. Although it makes a good use of all the information in the image, it is not the best way to represent the semantic meaning of the video sequences. According to the property of human perception, because of the selectivity of receiving and processing of visual information, important components in the image are usually paid more attention to, while those of less importance are ignored to various degrees. Moreover, processing the whole frame implies a higher computational complexity, due to more unnecessary information that has negative influences on the results and other following processes. In addition, clustering usually cannot be implemented until all the frames in the video sequences are analyzed, resulting in some additional time cost for calculation and large memory for data storage. To solve these problems, a visual attention model based method using online clustering is proposed in this paper. In Section 2, an overview of the proposed scheme and the whole algorithm flow are described in detail. Experimental results and discussions are given in Section 3. Section 4 concludes the whole paper.

2. Proposed video abstraction algorithm

Most of the existing video abstraction schemes perform feature extraction on all regions in each frame and treat all regions equally during feature comparison, and perform the clustering operation until all frames are ready, resulting in high complexity and large memory

Download English Version:

https://daneshyari.com/en/article/536999

Download Persian Version:

https://daneshyari.com/article/536999

Daneshyari.com