Ecological Informatics 23 (2014) 107–117

Contents lists available at ScienceDirect

Ecological Informatics

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

Shot based keyframe extraction for ecological video indexing and retrieval

G.G. Lakshmi Priya *, S. Domnic

Department of Computer Applications, National Institute of Technology, Tiruchirappalli 620015, INDIA

ARTICLE INFO

Article history:
Received 30 January 2013
Received in revised form 17 July 2013
Accepted 16 September 2013
Available online 29 September 2013

Keywords:
Video shot boundary detection
Keyframe extraction
ICSA
Ecological video summarization
Indexing and retrieval
Evaluation criteria

ABSTRACT

Among possible research area in multimedia, keyframe extraction is an important topic that provides video summarization, faster browsing and accessing of wide video collections. In this paper, we propose a new automatic shot based keyframe extraction for video indexing and retrieval applications. Initially, the frames are sequentially clustered into shots by using feature extraction, continuity value construction steps of shot boundary detection process and the shot frame clustering technique. The cluster having a larger dispersion rate is selected for inter cluster similarity analysis (ICSA) and the sub-shot based keyframes are extracted using ICSA. The proposed shot boundary detection algorithm and video keyframe extraction technique are implemented and evaluated on publicly available ecological video datasets. Compared with existing related algorithms, our method yields better F1-score of 94.2% for shot boundary detection and better results for keyframe extraction. The keyframes extracted by the proposed method are used for video indexing and retrieval.

1. Introduction

With the recent advancement in technologies, multimedia data are created easily which lead to the enormous availability of data on the web or in personal and professional databases. Due to this huge amount of data more specifically videos, the problem arises in accessing them. So, there is a need for an efficient tool to manage, index, access and retrieve the available information. Wider categories of professional people show their interest in different genre of information. One way for manipulating the video is the production of video summary or video abstract. Video summary is a compact representation of the video sequences by removing the duplicate data and is useful for various applications like video browsing, indexing and retrieval (Truong and Venkatesh, 2007). In general, video summary can be a preview of the video along with audio (video skimming) or set of keyframes chosen properly which depicts the visual content of the video sequence. This keyframe based summary is the simplest method for representing the video, which contains a set of motionless frames taken from shots and this compact representation can be used for video indexing. The use of these keyframes reduces the amount of information required for video indexing. By this, users can quickly browse over the video by viewing only a few highlighted keyframes.

Usually, keyframe extraction can be done by considering any of the two ways: a) assume the video file is already segmented into shots or b) segment the video into shots by the shot boundary detection technique. In the first case, depending on the nature of the shots (static/dynamic) (Amiri and Fathy, 2010), keyframes are extracted. A segmented shot is said to be static/dynamic based on the motion activity involved in the video sequences. However, more keyframes are necessary for representing the dynamic shots than the static ones. While considering on the second case, the video sequences are segmented into shots by the shot boundary detection method and later, the procedure for keyframe extractions is performed. The process of automatically segmenting the video into shots is shot boundary detection (SBD) or temporal video segmentation. A shot is defined as a sequence of frames captured by a single camera in an uninterrupted space and time (Yeo and Liu, 1995). Shots are the basic structural building blocks of the video. SBD is an important task in managing the video information for indexing, browsing, summarization and other content based operations. A SBD algorithm that detects the shot boundary more accurately is employed which in turn paves way for efficient keyframe extraction.

The main objective of our work is to propose a method for keyframe extraction of ecological videos which includes efficient shot boundary detection. In this paper, we propose a new: a) edge based feature for shot boundary detection, b) similarity measures to find the relationship between the consecutive frames, c) shot frame clustering procedure for segmenting the shots, and d) keyframe extraction for visual content representation of the video sequences. The evaluation is performed for both shot boundary detection and keyframe extraction with recent related works. We also use our keyframes in the work (Jiang et al., 2010), one of the participants of TRECVID semantic video indexing task (VIREO-374). In this paper, the significance of using multiple (varying number) keyframes per shot, instead of using single frame per shot in video retrieval is also studied.

The remainder of the paper is organized as follows: Section 2 reviews the existing works. In Section 3, the proposed method includes

* Corresponding author. Tel.: +91 9486322772.
E-mail address: gg_lakshmipriya@yahoo.co.in (G.G. Lakshmi Priya).

1574-9541/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecoinf.2013.09.003

feature extraction, similarity measures, shot frame clustering and keyframe extractions are presented in detail. Experimental evaluation of the proposed shot boundary detection and keyframe extraction are discussed in Section 4 and conclusions are derived in Section 5.

2. Literature review

Early works on keyframe extraction are based on methods: sampling based, segment based and shot based. In the sampling based method, random frames are selected as keyframe from uniformly sampled video. In segment based approach, the similar frames are grouped into clusters and keyframes are extracted from the clusters. Whereas in shot based approaches, after the video is segmented into shots, keyframes are extracted in various ways. In this section, we discuss the works related to the shot based keyframe extraction method.

2.1. Shot boundary detection

Until now, various methods are available in the literature for the identification of shot boundaries. In some of these methods (Boreckzy and Rowe, 1996; Lienhart, 1999), an abrupt transition is identified when a dissimilarity difference between successive frames surpasses a threshold. The difference measure is calculated at either pixel level or the block level. There exists various methods which include pixel-based features (Lienhart, 1999), color histogram (Gargi et al., 2001; Mas and Fernandez, 2003), edge based features (Lakshmi Priya and Domnic, 2012a,b; Yoo et al., 2006; Zabih et al., 1995), visual descriptors (Lee et al., 2003), Gabor texture (Barbu, 2009), multiple feature based (Lian, 2011),and the model based method (Mohanta et al., 2012). The two main factors which may mislead the detection process are object/camera motion and lighting effects.

In general, on comparing some of the low level features like pixel-based, color histogram, edge, and intensity; edge based features are robust to lighting effects and slightly robust to motion. In paper (Zabih et al., 1995), the algorithms based on color histograms, chromatic scaling and edge detection are compared and it was reported that the edge based approach is more accurate than histograms in detecting abrupt and is much less sensitive to motion than chromatic scaling. However, the histogram based method (Mas and Fernandez, 2003) is robust to fast motion but lacks their performance, when two consecutive frames have quite same histogram, while their contents are extremely dissimilar.

The detection of gradual transition is quite difficult compared to that of the abrupt transitions. Many researchers showed their interest in providing better detection for gradual transitions. In paper (Yoo et al., 2006), the authors have considered the variance distribution of edge information in the frame sequence for identifying the gradual transitions and used normalized correlation coefficient for identifying the occurrence of abrupt transitions in the video sequences. The main drawback of this method is that they are sensitive to camera, object motion and extensive content change within the shot. The authors (Adjeroh et al., 2009) have studied the problem of video shot boundary detection using an adaptive edge-oriented framework, where multiple multilevel features are used. The multiple features based shot boundary detection technique (Lian, 2011) was proposed, which uses predefined threshold value for each stage of the detection process. In model based approach (Mohanta et al., 2012), local and global features are used to represent frames. This method detects the abrupt and gradual transitions. There is a need for an efficient SBD technique that yields a maximum detection rate by suppressing the above discussed issues. The above mentioned works focus only on shot boundary detection task.

2.2. Keyframe extraction

Initially, with low level feature, keyframes are extracted by employing traditional clustering algorithms to select a number of frames as keyframes. A sequential clustering algorithm is proposed (Zhuang et al., 1998), where new clusters are created sequentially once the dissimilarity to previous key-frame reaches above a specified threshold. From those 'big' clusters, frames closer to the cluster centroids are selected as key-frames. In the paper (Girgensohn and Boreczky, 2000), the position of the keyframes in time series is also used in the clustering process and hierarchical clustering reduction is performed, which yields different levels of abstraction. There are some clustering methods (Zhuang et al., 1998) that require user intervention in setting the maximum cluster size in the beginning of the process. The frames that are closest to the centroids of each cluster are extracted as keyframes. In the paper (Gong and Liu, 2000), the video is summarized with a clustering algorithm based on single value decomposition (SVD). The refined feature space obtained by the SVD is clustered and a keyframe is extracted from each cluster. The above discussed methods focus mainly on keyframe extraction irrespective of segmenting the shots.

In the papers (Borth et al., 2008; Cernekova et al., 2006; Gianluigi and Raimondo, 2006), the authors proposed shot based keyframes extractions for generating video summaries. Cernekova (Cernekova et al., 2006) proposed a method for detecting shot boundaries in video sequences and for extracting keyframes using information theory based metrics. In the paper (Gianluigi and Raimondo, 2006), an approach for the selection of keyframes that determines the complexity of the sequence in terms of changes in the pictorial content using three visual features: its color histogram, wavelet statistics and an edge direction. Borth uses the shot boundary detection to segment the video into shots and the k-means algorithm to determine cluster representatives for each shot that are used as keyframes (Borth et al., 2008). In the paper (Besiris et al., 2009), the author proposed an automatic video summarization technique based on graph theory methodology and the dominant sets clustering algorithm. Amiri (Amiri and Fathy, 2010) proposed a Hierarchical keyframe based video summarization system using QR-decomposition. Based on the dynamicity of the shots the number of keyframes is selected.

Some works have been proposed for video keyframe extraction in the domain of entertainment (Ying, 2006), education (Choudary and Liu, 2007), sport events (D'Orazio and Leo, 2010) and wildlife (Yong et al., 2011). For example, biologists struggle towards understanding the behavior of plants, insects and animals, and face more problems in observing and analyzing the natural environment. To address this problem, more researchers showed their interest in analyzing the ecological data. Semantically integrated data models have been proposed for species observation (Veen et al., 2012). In the paper (Groom et al., 2013), the authors have analyzed the distribution of marine bird from the remote sensing image data. Also, to help the ecologist study the animal behavior, the authors have developed an automatic tool to visualize the animals' movements from satellite collars and radio-telemetry (Kavathekar et al., 2013).

Regarding wildlife video analysis specifically, animal hunt detection was proposed by Haering et al (Haering et al., 2000). Their event-based method utilizes a combination of color, texture and motion features. In another noticeable work, a clustering-based approach is proposed for wildlife video abstraction using Gaussian mixture models (GMM) (Gibson et al., 2002). Working with wildlife videos, the framework in the paper (Yong et al., 2011) undergoes image segmentation, feature extraction and matching of image blocks and then a co-occurrence matrix of semantic labels is constructed to represent the semantic context within the scene.

A video representation (summary) is considered to be good, if the set of keyframes effectively represents the visual content of the video sequence. There are no benchmarking or ground truth results for keyframe extraction algorithms. The measures used for video summarization can be used for keyframe extraction. To measure the effectiveness of the video summary (Truong and Venkatesh, 2007), the existing methods are grouped as descriptive evaluation, objective metrics and