Contents lists available at ScienceDirect

Information Processing and Management

What do you wish to see? A summarization system for movies based on user preferences

Rajkumar Kannan [a,*], Gheorghita Ghinea [b], Sridhar Swaminathan [c]

[a] King Faisal University, Al Ahsa 31982, Saudi Arabia
[b] Brunel University, Middlesex UB8 3PH, United Kingdom
[c] Bishop Heber College (Autonomous), Tiruchirappalli 620017, India

ARTICLE INFO

ABSTRACT

Video summarization aims at producing a compact version of a full-length video while preserving the significant content of the original video. Movie summarization condenses a full-length movie into a summary that still retains the most significant and interesting content of the original movie. In the past, several movie summarization systems have been proposed to generate a movie summary based on low-level video features such as *color*, *motion*, *texture*, etc. However, a generic summary, which is common to everyone and is produced based only on low-level video features will not satisfy every user. As users' preferences for the summary differ vastly for the same movie, there is a need for a personalized movie summarization system nowadays. To address this demand, this paper proposes a novel system to generate semantically meaningful video summaries for the same movie, which are tailored to the preferences and interests of a user. For a given movie, shots and scenes are automatically detected and their high-level features are semi-automatically annotated. Preferences over high-level movie features are explicitly collected from the user using a query interface. The user preferences are generated by means of a stored-query. Movie summaries are generated at shot level and scene level, where shots or scenes are selected for summary skim based on the similarity measured between shots and scenes, and the user's preferences. The proposed movie summarization system is evaluated subjectively using a sample of 20 subjects with eight movies in the English language. The quality of the generated summaries is assessed by *informativeness*, *enjoyability*, *relevance*, and *acceptance* metrics and *Quality of Perception* measures. Further, the usability of the proposed summarization system is subjectively evaluated by conducting a questionnaire survey. The experimental results on the performance of the proposed movie summarization approach show the potential of the proposed system.

1. Introduction

The past decade has witnessed an explosive growth of digital videos, both over the Internet and on home computers. However, browsing through lengthy and voluminous video collection becomes tedious to the user, if the videos have little or no relevant content. Moreover, searching for interesting segments within the videos is time consuming (Parshin & Chen, 2004). Video summarization systems generate a compact version of a lengthy video and help the users to watch its

significant segments (Money & Agius, 2008; Tsai, Kang, Lin, & Lin, 2013). According to the types of content used for video analysis, existing video summarization methods can be classified into *cognitive-level* approaches and *affective-level* approaches (Tsai et al., 2013). The cognitive-level video summarization approaches (Evangelopoulos et al., 2013; Ma, Lu, Zhang, & Li, 2002; Ngo, Ma, & Zhang, 2005; Taskiran, Pizlo, Amir, Ponceleon, & Delp, 2006; You, Liu, Sun, & Li, 2007) extract low-level features such as *color, motion, texture, and audio*, *visual* and *textual* saliencies to identify important segments of a video. On the other hand, the affective-level summarization approaches (Joho, Jose, Valenti, & Sebe, 2009; Katti, Yadati, Kankanhalli, & Tat-Seng, 2011; Peng et al., 2011) generate video summaries by modeling the affective video content by exploiting users' feedbacks/responses while watching a video. In general, most of these approaches extract significant segments of a video based on low-level features that are considered to get the users' attention. However users always desire video summaries that are generated based on the high-level features such as events, semantic concepts etc., rather than low-level features alone (Lie & Hsu, 2008). It is rather unsurprising, therefore that the well-known semantic gap problem is thus shown to also exist in video summarization.

Most of the existing video summarization systems are generic in nature. That is, for a video, these systems create a video summary that is common for all the users. Generic video summarization will not be sufficient when the users' needs and interests differ and change over a time (Lie & Hsu, 2008). Thus, the users are seldom satisfied by a *generic video summary* (also called *non-tailored video summary*) produced by a video summarization system (Lie & Hsu, 2008). This is because the produced video summary may not contain video content of the particular event, semantic concept or genre liked by the user. A generic video summarization system that produces video summaries based only on low-level video features cannot process a semantic level query from a user, such as '*summarize all the action events of a movie*'. Personalized or tailored video summarization is a useful technique for producing tailored video summaries to the users based on their needs and interests (Ghinea, Kannan, Swaminathan, & Kannaiyan, 2014; Lie & Hsu, 2008). Also, recent information retrieval and filtering systems have started tailoring or personalizing the results by adjusting to individual user's needs and interests. Accordingly, we argue that the criterion used to summarize a video should be the user's preferences and interests over the high-level features of a video.

Film is an art form that offers a practical, environmental, pictorial, dramatic, narrative and musical medium to convey a story (Monaco, 2000). Trailers, which are short summaries of movies, have been used for decades to promote movies. Creating movie summaries manually by domain experts is a tedious and time consuming task as the users' viewing time constraints and preferences change over a time. Movie summarization, being a special class of video summarization, is particularly challenging since a large variety of movie scenarios and film styles complicate the summarization problem (Tsai et al., 2013). Movie summaries help the user to decide whether to watch an entire movie or not. Most of the movie summarization methodologies generate summaries based on the visual attention and motion activities in a movie (Evangelopoulos et al., 2013; Ngo et al., 2005). These summaries will be semantically meaningful only if the summarization methodology considers high-level movie features (Lu, Lyu, & King, 2005). Since characters are considered as the important high-level features of a movie, recent approaches for movie summarization (Sang & Xu, 2010; Tsai et al., 2013; Weng, Chu, & Wu, 2009) identify roles in a movie by exploiting social network analysis and role recognition. Apart from characters, users might also be interested in important events, and semantic concepts in a movie for summarization.

Movie videos contain rich high-level features such as *characters*, *events*, *semantic concepts* and *spoken content*. The flourishing movie industries produce more than 4500 movies every year (Sang & Xu, 2010). Manual annotation of the aforementioned movie features for a large number of movie videos is tedious, time consuming and laborious. This necessitates movie summarization approaches to exploit efficient indexing and automatic annotation techniques for pre-processing. Video Content Analysis (VCA) greatly helps the users for an effective media management including indexing, retrieval, and summarization. Bridging the "semantic gap" has always been one of the notorious and biggest challenges in video content management, i.e. allowing the users to browse, retrieve, and summarize video content at semantic level (Li, Errico, Pan, & Sezan, 2004). Even though movie content analysis does not always need a real-time processing as required for surveillance, and sports video analysis, the rapid growth of movie videos necessitates an efficient movie content analysis. Thanks to the recent advancements in the fields of Computer Vision, and Pattern Recognition in high-level feature detection, and recognition which has let the Multimedia Information Retrieval (MIR) applications to exploit faster and accurate automatic annotation techniques.

As shown in Fig. 1, the granularity of movies goes from video frames, shots to scenes and substories. A video should be broken down into a set of segments which are either shots or scenes for efficient indexing. *Shot Boundary Detection* (SBD) (Smeaton, Over, & Doherty, 2010) and *Scene Change Detection* (SCD) (Sidiropoulos et al., 2011) techniques temporally decompose a video into more manageable units. Also the manual efforts for annotation of shots and scenes can be reduced by automatic video content analysis techniques. Moreover, *face recognition* (Turk & Pentland, 1991) and *large scale semantic concept detection* (Chao, 2012) techniques can be utilized for annotating visual features of a movie automatically. The spoken content of a movie can also be automatically annotated using *Automatic Speech Recognition* (ASR) techniques, or *closed captions* that are available in the web as subtitles. In personalized movie summarization, users may be interested in all possible high-level movie features. User profiles, as employed in recommender systems, can be utilized for such a personalization task.

Unlike most of the generic video summarization approaches which summarizes a video based on the objective criterion such as saliencies, user feedbacks/responses, character analysis, information coverage, and diversity, the proposed summarization methodology generates semantically meaningful personalized movie summaries by exploiting user's subjective preferences over the high-level visual and textual features of a movie. The novelty of the proposed approach as opposed