

Chapter 13

Content Coverage and Redundancy Removal in Video Summarization

Hrishikesh Bhaumik

RCC Institute of Information Technology, India

Siddhartha Bhattacharyya

RCC Institute of Information Technology, India

Susanta Chakraborty

Indian Institute of Engineering Science and Technology, India

ABSTRACT

Over the past decade, research in the field of Content-Based Video Retrieval Systems (CBVRS) has attracted much attention as it encompasses processing of all the other media types i.e. text, image and audio. Video summarization is one of the most important applications as it potentially enables efficient and faster browsing of large video collections. A concise version of the video is often required due to constraints in viewing time, storage, communication bandwidth as well as power. Thus, the task of video summarization is to effectively extract the most important portions of the video, without sacrificing the semantic information in it. The results of video summarization can be used in many CBVRS applications like semantic indexing, video surveillance copied video detection etc. However, the quality of the summarization task depends on two basic aspects: content coverage and redundancy removal. These two aspects are both important and contradictory to each other. This chapter aims to provide an insight into the state-of-the-art approaches used for this booming field of research.

1. INTRODUCTION

With the ever decreasing cost of digital storage devices and advancement of technology in recent years, video recorders have gained immense popularity. A large number of videos are produced and uploaded at an ever increasing rate. Digital libraries and video repositories like YouTube, DailyMotion, MyVideo etc. enable users to upload, retrieve, view and share videos over the internet. This has also been facilitated

DOI: 10.4018/978-1-5225-0498-6.ch013

by a many fold increase in communication bandwidth provided by Internet Service Providers (ISPs). Due to its inherent structure, a video encompasses the other three media types as well, i.e. text, image and audio, combining them into a single data stream. As a consequence, the analysis and retrieval of videos has attracted much attention from researchers and application developers. The research issues include feature extraction, similarity/dissimilarity measures, segmentation (temporal and semantic), key-frame extraction, indexing, annotation, classification and retrieval of videos. Video summarization is an application which lies on the crossroads of these research issues. The objective of video summarization is to reproduce and represent a video to the user in a concise manner such that its overall semantic meaning is preserved. The time and bandwidth constraints besetting the user are balanced to a great extent by such summarization task. Real-time video summarization helps in immediate indexing of videos for facilitating retrieval from the repository. Also it facilitates the user in assessing whether watching the entire video would be useful or not. Video summarization and representation through key frames have been frequently adopted as an efficient method to preserve the overall contents of the video with a minimum amount of time. The types of video summarization include static (also known as storyboard representation) and dynamic (also called video skimming). Storyboard or static representation is produced by means of selecting key-frames from a set of video frames, obtained after decomposing a video. Audio clips have temporal characteristics and are therefore not an integral part of static summaries. However, some keywords in the form of meta-data may be tagged with the static video frames to assist in keyword based indexing and retrieval. The set of key-frames serve as features of the video and these key-frames help in indexing task. During the process of retrieval, the key-frames corresponding to a video may be matched with the key-frames extracted from the query video. As such, the task of video mining is facilitated through this process.

On the contrary, video skimming techniques aim at selecting shots or scenes from the entire video and collating it together to produce a meaningful summary. Due to the intrinsic nature of dynamic summaries, the other three media types may feature in it. This enables the user to assimilate the contents of the video in a short time, so as to ascertain if the same is of interest to him/her. Video summarization is one of the key application areas of video indexing and content based video retrieval. Video summarization and representation through key frames have been frequently adopted as an efficient method to preserve the overall contents of the video with a minimum amount of time. The results of video summarization can be reused in many areas such as content-based video retrieval (Patel & Meshram, 2012), semantic indexing (Papadimitriou et al., 1998; Marcus & Maletic, 2003; Smith et al., 2003), Copied Video Detection (CVD) (Kim et al., 2008) etc. A video summary can be categorized as a static summary (storyboard), which is a set of selected key-frames (Chang et al., 1999; DeMenthon et al., 1998; Hanjalic & Zhang, 1999; Yeung & Yeo, 1997), or a dynamic video representation (Lienhart, 2000; Nam & Tewfik, 1999; Gon & Liu, 2000a; Smith & Kanade, 1997) formed by concatenating short video segments. Dynamic video summarization methods include semantic analysis (Nam & Tewfik, 1999; Ma et al., 2002), motion model (Ma & Zhang, 2002; Vasconcelos & Lippman, 1998), expectation maximization (Orriols & Binefa, 2001) and singular value decomposition (Gong & Liu, 2000a). As there are plentiful redundancies in the frames of a shot, the frames which appropriately portray the contents in the best possible way are selected (Sze et al., 2005; Truong & Venkatesh, 2007; Besiris, 2007; Mukherjee et al., 2007). These frames are termed as key-frames. The various features taken into consideration for key-frame extraction include edges of objects in the frames, optical flow, RGB histograms, shapes of objects, spatial distribution of MPEG-7 (Narasimha et al., 2003) discrete cosine coefficient of MPEG (Moving Picture Experts Group) motion vectors (Wang et al., 2007), and derived features due to camera movement (Guironnet

21 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/content-coverage-and-redundancy-removal-in-video-summarization/159443

Related Content

Transforming the Classroom Experience Through Transhumanism: Education as the Learning Organization

Mary Beth Klinger and Teresa L. Coffman (2019). *Handbook of Research on Learning in the Age of Transhumanism* (pp. 134-156).

www.irma-international.org/chapter/transforming-the-classroom-experience-through-transhumanism/227909

CoAP-Based Lightweight Interoperability Semantic Sensor and Actuator Ontology for IoT Ecosystem

Sukhavasi Suman, Thinagaran Perumal, Norwati Mustapha, Razali Yaakob, Mohd Anuaruddin Bin Ahmadon and Shingo Yamaguchi (2021). *International Journal of Ambient Computing and Intelligence* (pp. 92-110).

www.irma-international.org/article/coap-based-lightweight-interoperability-semantic-sensor-and-actuator-ontology-for-iot-ecosystem/275760

Exploring Video Sharing Websites Content with Machine Learning

Nan Zhao, L  c Baud and Patrick Bellot (2017). *Artificial Intelligence: Concepts, Methodologies, Tools, and Applications* (pp. 1770-1789).

www.irma-international.org/chapter/exploring-video-sharing-websites-content-with-machine-learning/173402

A Framework for Assessing Enterprise-Wide SOA Implementation Readiness

Mahshidah Abdul-Manan and Peter Hyland (2013). *International Journal of Intelligent Information Technologies* (pp. 21-37).

www.irma-international.org/article/framework-assessing-enterprise-wide-soa/77872

Blockchain Technology and Additive Manufacturing

Yesim Can Saglam (2023). *Streamlining Organizational Processes Through AI, IoT, Blockchain, and Virtual Environments* (pp. 85-105).

www.irma-international.org/chapter/blockchain-technology-and-additive-manufacturing/325338