Online Video Summarization Based on Local Features

Javier Iparraguirre, Electronics Engineering Department, Universidad Tecnológica Nacional and Images Science Laboratory, Universidad Nacional del Sur, Bahía Blanca, Argentina

Claudio A. Delrieux, Images Science Laboratory, Universidad Nacional del Sur, Bahía Blanca, Argentina

ABSTRACT

Digital video has become a very popular media in several contexts, with an ever expanding horizon of applications and uses. Thus, the amount of available video information is growing almost limitless. For this reason, video summarization continues to attract the attention of a wide spectrum of research efforts. In this work, the authors present a novel video summarization technique based on tracking local features among consecutive frames. The authors’ approach operates on the uncompressed domain, and requires only a small set of consecutive frames to perform. Additionally, the authors’ algorithm processes the video stream directly and produces results on the fly. The authors tested their implementation with standard available datasets, and compared the results with the most recently published work in the field. The achieved results are similar or better in quality than the best performing proposals, with the additional advantage of being able to process the stream directly in the uncompressed domain.

Keywords: Computer Vision, Keyframe Selection, Local Features, Video Abstraction, Video Skimming, Video Summarization

INTRODUCTION

In recent years, digital video became a vastly used media in several contexts and applications. The generation and availability of digital video from different sources is growing at an exponential rate (Money & Agius, 2008). This poses several challenges for users in order to manipulate a vast collection of videos. One major way to simplify and accelerate the access to a particular information item in a video sequence is to provide abridged (albeit complete in some sense) representations of the whole content. This method saves the users the burden of having to watch complete videos to decide whether and where the required information is present. Video summarization aims to provide these condensed versions in a consistent and predictable way.

Summarization techniques must produce an intelligible output that can be useful to human users. There are multiple aspects to consider in the manipulation of digital video. On the one hand, any processing must consider the capture, encoding, and compression techniques that are applied in the digital medium. On the other hand, the psychological features of the human perceptual system should be taken into account for an adequate processing and manipulation.

DOI: 10.4018/ijmdem.2014040103
This renders video summarization as a very complex and difficult task to assess.

Video summarization can be classified into two main categories (Truong & Venkatesh, 2007). The first one is keyframe-based summarization. The output of this process is a storyboard or a static summarization. The second case is sequence-based summarization, which produces a short version of the original material (called video skim in the literature). In short, these two approaches are referred to as storyboarding and skimming, respectively.

Another relevant feature to consider is whether the summarization method requires the complete video sequence in order to proceed (batch summarization), or if it may perform directly over the video stream, which makes it adequate for online processing. Finally, depending on the way the video information is accessed, it is possible to apply the processing in the compressed or in the uncompressed domain. Compressed domain techniques employ some of the features that are provided by the existing video encoders. By contrast, uncompressed summarization uses all the information available in the frames.

In this paper we present a novel technique that is adequate both for storyboarding and for skimming. The technique performs only with the local features of a small set of consecutive (uncompressed) frames. Hence, our method may operate directly with the video stream and, since it performs at an adequate speed, it can be applied to summarize video in real time. The working parameters of our method are just a few and have a very intuitive meaning, making them easy to tune adaptively to any working condition. One of the key components of our technique is based on the Speeded-up Robust Features (SURF) algorithm (Bay, Ess, Tuytelaars, & Van Gool, 2008).

We tested the behavior of our technique with a set of standard datasets. The resulting summarizations are equivalent in quality to the best published results in the field. In addition, our processing was performed directly on the uncompressed video stream, which makes it useful for live applications. Finally, most of the alternative summarization techniques available in the literature appear to require the fine tuning of a large set of parameters, while in our proposal just a couple of settings are required. To the best of our knowledge, this is so far the first on-line, uncompressed proposal published in the literature.

This work is a part of an ongoing research project (Iparraguirre & Delrieux, 2013). In this paper we publish a more thorough description of our methodology and we explain in detail the meaning and tuning of the working parameters of the algorithm. We also provide a more extensive discussion on the results. In the next section we introduce the relevant prior work related to video summarization. Also, for completeness, we describe in the basic ideas underlying the SURF algorithm in the third section. Afterwards, we introduce our method and present the most important implementation details. The results are presented in the fifth section. Finally, we discuss the conclusions and propose further work in the final section of this paper.

RELATED WORK AND CONCEPTS

As mentioned previously, among the purposes of video summarization we can consider to enable faster browsing of large video collections and efficient indexing and access. The result of such a system consists of an automatically generated short summary of a video, which can either be a static summary (a set of keyframes or storyboard) or a dynamic summary (a set of shots or skim). These reduced versions are generally produced by taking into account some similarities or domain-specific relationships among video frames or shots. A general review of video summarization can be found in Truong and Venkatesh (2007) or in Money and Agius (2008).

Research on this topic has been very active recently, and therefore there is a good number of proposed techniques available. The earliest proposals followed variations of a similar approach, which starts creating a large matrix
Methods for Dealing with Dynamic Visual Data in Collaborative Applications - A Survey
[www.igi-global.com/chapter/methods-dealing-dynamic-visual-data/27039?camid=4v1a](www.igi-global.com/chapter/methods-dealing-dynamic-visual-data/27039?camid=4v1a)

QoS Routing for Multimedia Communication over Wireless Mobile Ad Hoc Networks: A Survey
[www.igi-global.com/article/qos-routing-for-multimedia-communication-over-wireless-mobile-ad-hoc-networks/176640?camid=4v1a](www.igi-global.com/article/qos-routing-for-multimedia-communication-over-wireless-mobile-ad-hoc-networks/176640?camid=4v1a)