Video Summarization

Guillermo Camara, Orientador do Edward Jorge Yuri Cayllahua Cahuina
PPGCC - Programa de Pós-Graduação em Ciência da Computação
UFOP - Universidade Federal de Ouro Preto
Ouro Preto, Minas Gerais, Brasil
email: gcamarac@gmail.com, ecayllahua1@gmail.com

***Resumo*—In this paper we present several approaches for video summarization, we focus our efforts in static video summarization. Different methods are reviewed and we propose one method based on spatiotemporal features for its implementation. This method is proved to be effective in video summarization according to the literature.**

***Keywords*-Video Summarization, static video storyboard, algorithms, spatiotemporal features.**

I. Introduction

The volume of multimedia information such as text, audio, still images, animation, and video is growing every day. The accumulated volume of this information can become a large collection of data, it would be an arduous work if a human tries to process such a large volume of data and even, at a certain scale, it would be impossible. A perfect example of this is video.

Video information is growing exponentially, each day an enormous quantity of video is uploaded to the internet, TV video information is generated every day, security cameras generate hours of video. It is necessary to develop a model in order to manage all this information. Video summarization aims to give a user a synthetic and useful visual summary of a video sequence.

Thus, a video summary is a short version of an entire video sequence. The video summary can be represented in two fashions: a static video storyboard and a dynamic video skimming. Dynamic video skimming consists in selecting the most relevant small dynamic portions of audio and video in order to generate the video summary. On the other hand, static video storyboard is interested in selecting the most relevant frames of a video sequence and generate a generate the correspondent video summary. Obviously, the key part is to recognize these relevant frames or portions of video, and this adds a certain subjectivism to the methods in the literature because different methods have different points of view of what is relevant and what is not.

Many methods have been proposed for video summarization, dynamic video skimming usually have very complex models which demands a long time for its implementation, so for this first part we have chosen to develop a static video storyboard which is more suitable for the time frame we have been given.

We plan to implement the method proposed by [1]. This method first extracts the frames from the video and computes their spatiotemporal features. For this extraction the method uses the spatiotemporal Hessian matrix which proves to be a good feature extractor and also provides a measure of the activity that happens within each frame. Later, this information is processed to extract the most important frames (*keyframes*) based in the frames with higher activity, after that it performs a cleaning process where junk frames are deleted and finally it constructs a clip with the *keyframes* that are still relevant. According to their experiments this method proves to be effective.

Other methods such as the one by [2] rely on machine learning, detecting what they consider relevant features to make their model more robust to noises in the video. In [3] a greedy algorithm is proposed which has more simplicity and a good performance but is not as robust as [2]. Then in [4] they present an algorithm based on dynamic programming where they use a MINMAX optimization to detect the *keyframes* although we don't know how robust their model is.

This article is structured as follows. Section II presents the justification of our topic, outlining its importance. The main goals we pursue in this article are presented in Section III. Section IV will explain how we plan to develop our proposition and finally in Section V we list the results we expect to achieve.

II. Justification

In order to process the enormous volume of video information created every day, effective methods for video summarization need to be created. This is the reason why during the last years video summarization has been an interesting topic for many researchers and is still an open problem. Video summarization has a direct impact on commercial applications where most of this video information is usually created.

III. Goals

The main goal we want to achieve is to develop a method based on [1] for static video summarization.

IV. Methodology

In order to achieve our goals, we intend to follow the next steps:

- bibliography revision. Until $08/06/2011$.
- Implementation. Until $24/06/2011$. The whole implementation based on the method proposed by [1] and we will evaluate the possible implementation of [3] to run comparisons.
- Run tests on our implementation(s).

V. Expected Results

At the end of this work, it is expected to have a method that effectively summarizes videos, although it may still not be that robust when dealing with noises in the video. We will focus on this noises and a more robust model in future work.

Referências

[1] R. Laganière, P. Lambert, and B. E. Ionescu, "Video summarization from spatio-temporal features," *Proceedings of the 2nd ACM Workshop on Video Summarization, ACM*, 2008.

[2] W. Ren and Y. Zhu, "A video summarization approach based on machine learning," *International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IEEE*, 2008.

[3] L. Coelho, L. A. D. S. Cruz, L. Ferreira, and P. A. Assunção, "An improved sub-optimal video summarization algorithm," *International Symposium ELMAR (Electronics in Marine) - 2010, IEEE*, 2010.

[4] Z. Li, G. M. Schuster, and A. K. Katsaggelos, "Minmax optimal video summarization," *IEEE Transactions On Circuits And Systems For Video Technology*, 2005.