Event detection in wedding videos

Overview

We have developed a system for key event extraction from and summarization of video that does not rely on annotated training data, but uses textual and visual data. The model is applied on videos that document royal weddings resulting in valuable visualizations of the key events in the video.

In depth description

In a first step we have learned important wedding concepts in Wikipedia articles about weddings as frequently occurring words and phrases that signal events in these articles. We used a state-of-the-art event recognizer developed at KU Leuven. Given videos of weddings we first detect the key events in the textual data of the video (the raw speech transcripts and the subtitles were used). The names and faces alignment tool was then applied to identify wedding events that involve key persons in the video. This step allows indexing selected frames from the video with key wedding events and their participants. In a next step a best frame is selected among the frames that represent a specific event based on a k-medoid clustering algorithm of the frames. These best frames form the video summary. We have demonstrated and evaluated this task on a dataset of Britain's royal weddings. We compared two settings, one using the transcripts and one using the subtitles. We are currently performing a complete evaluation of the system.

Potential fields of application

This technology can be included as a part of a media search engine. It provides an unsupervised way to dtect video concepts based on textual and visual data.

Possibilities for exploitation

They include the further development of the technology within the framework of the iV&L EU Cost action (IC1307).

Further Information

Further technical information is available in TOSCA-MP deliverable D3.4. A journal publication is in preparation and will be submitted soon.

Contact person

Prof. Marie-Francine Moens
Department of Computer Science
Celestijnenlaan 200A
B-3001 Heverlee, BELGIUM
sien.moens@cs.kuleuven.be