

Visual and multimodal linking

Overview

A software module for determining links between video segments based on visual and textual similarity.

In depth description

The software component determines links between segments of a video collection based on visual and textual similarity, e.g., related items to a search result.

Visual similarity is based on extracting and matching SIFT local descriptors of the videos in the collection. Similarity includes duplicates and near duplicates, as well as partial matches (same background, picture-in-picture). The results are pairs of matching video segments together with a similarity score. Both feature extraction and matching are implemented on graphics processors (GPU) for speedup.

Text similarity requires the availability of text resources such as closed captions or results of automatic speech recognition. Concepts and named entities are detected and classified using linked data resources (e.g., DBpedia, Freebase, Geonames), and different scores are assigned to matching terms based on their categories.

The final output is a list of segments ranked by similarity. The combination of visual and text-based linking exploits the complementarities between the two types of information.

Potential fields of Application

The component can be used for organising medium sized content collection (several hundreds of hours) by creating links between video segments based on visual or topical similarity. This includes organising production-size data collections, smaller video archives, presentation of search results, video collections of specific events, etc.

Possibilities for exploitation

The software component can be licensed to interested parties.

Further Information

An overview of the method can be found in http://ceur-ws.org/Vol-1043/mediaeval2013_submission_70.pdf

Contact Person

Georg Thallinger
JOANNEUM RESEARCH
DIGITAL – Institute for Information and Communication Technologies
Steyrergasse 17
8010 Graz
Austria
georg.thallinger@joanneum.at