Hessisches Kompetenzzentrum für Hochleistungsrechnen

Image and Scene Search in Historical Collections of DDR Television Recordings in the German Broadcasting Archive (DRA) by Automatic Content-based Video Analysis



Project Manager
Dr. Markus Mühling

Principal Investigator
Prof. Dr. Bernd Freisleben

Project Term
2013 - 2016

Project Areas
Social and Cultural Anthropology, Non-European Cultures, Jewish Studies and Religious Studies

Clusters
MaRC2 Cluster Marburg

Institute
Fachbereich Mathematik und Informatik

University
University of Marburg

Introduction

The rapidly increasing proliferation of digital images and videos leads to a situation where content-based search in multimedia databases becomes more and more important. A prerequisite for effective image and video search is to analyze and index media content automatically. Due to the large amount of data and the associated hardware requirements, especially in the field of video retrieval, high performance computing plays an important role in this context. The high performance cluster MaRC has been used to support research in a collaborative project funded by German Research Foundation (DFG) called „Image and Scene Search in Historical Collections of DDR Television Recordings in the German Broadcasting Archive (DRA) by Automatic Content-based Video Analysis“. The goal of this project is to allow users to search within the DDR television recordings based on innovative methods from the field of content-based image and video retrieval.

Methods

The developed algorithms bundled in the video content analysis toolkit „Videana“[1] provide methods for shot boundary detection, face detection, face recognition, text detection,

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg

optical character recognition, and visual concept detection. The visual concept detection approach is used to automatically assign semantic tags to images and videos. The resulting index can be used to efficiently process search queries in large-scale multimedia databases by mapping query terms to semantic concepts. Thus, the semantic index serves as an intermediate description to bridge the „semantic gap“ between the data representation of images and videos and their human interpretation. State-of-the-art visual concept detection systems mainly rely on local image features. Similar to the representation of documents in the field of text retrieval, an image or video shot can then be represented as a bag of visual words by mapping the local descriptors to a visual vocabulary. In this context, we developed enhanced local descriptors with respect to the spatial extent, the integration of color, and the use of audio information.[4-6]

Furthermore, novel object-based features are used as additional inputs for the classifiers (i.e., support vector machines) used in our approach. Another issue in the field of visual concept detection is the difference in the visual appearance of semantic concepts across different domains, for example, news videos and user-generated YouTube videos. Adaptive methods have been investigated to improve the cross-domain capabilities of concept detectors.[2] Finally, a long-term incremental web-supervised concept learning approach [3] has been presented that exploits social web resources as training data and thus minimizes the human effort for building concept models. The evaluation of these novel approaches is based on large sets of images and videos, and often several hundreds of visual concepts had to be trained and tested. The MaRC cluster has been used for these extensive experiments, and significant performance improvements in terms of retrieval performance have been obtained. For the DDR television recordings, a concept lexicon of 106 visual concepts, such as „beach“, „winter“, „woman“, „airport“, or more special concepts such as „brotherly kiss“ and „Trabant“ has been developed. First evaluations on 220 hours of DDR television recordings show very promising results.

Outlook

Recently, artificial neural networks are experiencing a renaissance in computer vision, and we are currently investigating hybrid approaches using bag-of-visual words in combination with convolutional neural networks.

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg

Reference

[1] R. Ewerth, M. Mühling, T. Stadelmann, J. Gllavata, M. Grauer, and B. Freisleben (2008), Videana: A Software Tool for Scientific Film Studies. Digital Tools in Media Studies, Transcript Verlag Bielefeld: 101–116. https://doi.org/10.14361/9783839410233

[2] R. Ewerth, M. Mühling, and B. Freisleben (2012), Robust Video Content Analysis via Transductive Learning. ACM Transactions on Intelligent Systems and Technology: 1–26. https://doi.org/10.1145/2168752.2168755

[3] R. Ewerth, K. Ballafkir, M. Mühling, D. Seiler, and B. Freisleben (2012), Long-Term Incremental Web-Supervised Learning of Visual Concepts via Random Savannas. IEEE Transactions on Multimedia: 1008–1020. http://dx.doi.org/10.1109/TMM.2012.2186956

[4] M. Mühling, R. Ewerth, and B. Freisleben (2011), On the Spatial Extents of SIFT Descriptors for Visual Concept Detection. In Proc. of the 8th International Conference on Computer Vision Systems: 71–80.

[5] M. Mühling, R. Ewerth, J. Zhou, and B. Freisleben (2012), Multimodal Video Concept Detection via Bag of Auditory Words and Multiple Kernel Learning. In Proc. of the 18th International Conference on Multimedia Modeling: 40–50.

[6] M. Mühling (2014), PhD thesis: Visual Concept Detection in Images and Videos, University of Marburg, Germany https://doi.org/10.17192/z2014.0483

Last Update: 2022-06-17 21:24