Multimed Tools Appl (2016) 75:8969–8972
DOI 10.1007/s11042-016-3683-9

GUEST EDITORIAL

Guest Editorial: Content-based Multimedia Indexing

Harald Kosch[1] · Georges Quénot[2,3]

Published online: 20 July 2016

Content-Based Multimedia Indexing systems aim at providing easy, fast and accurate access to large multimedia repositories. Research in Content-Based Multimedia Indexing covers a wide spectrum of topics in content analysis, content description, content adaptation and content retrieval. Various tools and techniques from different fields such as Data Indexing, Machine Learning, Pattern Recognition, and Human Computer Interaction have contributed to the success of multimedia systems.

Although, there has been a significant progress in the field, we still face situations when the system shows limits in accuracy, generality and scalability. Hence, the goal of this special issue is to bring forward the recent advancements in content-based multimedia indexing. The papers included contain significant original new information and ideas.

For this special issue, we received a total of 14 submissions, of which 8 were accepted after a rigorous review process that consisted of several rounds of review. The 8 selected papers cover a wide range of problems in content indexing of multimedia data, including image, audio, video and multi-modal content.

The first paper "*A comparative study for multiple visual concepts detection in images and videos*" (DOI 10.1007/s11042-015-2730-2), co-authored by Abdelkader Hamadi, Philippe Mulhem and Georges Quénot, describes a comparative study and new methods for multi-concept detection in images and videos. The authors propose original fusion algorithms of one-concept detectors and propose a new stacking scheme for them. Large evaluations on the

✉ Georges Quénot
Georges.Quenot@imag.fr

Harald Kosch
harald.kosch@uni-passau.de

[1] University of Passau, Passau, Germany

[2] University Grenoble Alpes, LIG, F-38000 Grenoble, France

[3] CNRS, LIG, F-38000 Grenoble, France

PASCAL VOC'12 collection regarding the detection of pairs and triplets of concepts were done. In addition, the evaluation was extended to the TRECVid 2013 dataset for infrequent concept pairs' detection.

The second paper "*Naming multi-modal clusters to identify persons in TV Broadcast*" (DOI 10.1007/s11042-015-2723-1), co-authored by Johann Poignant, Guillaume Fortier, Laurent Besacier and Georges Quénot, brings forward a new method for identifying persons in TV Broadcasts by including written names during the diarization process. A multi-modal matrix of distances between speaker turns and face tracks is constructed, on this matrix, an agglomerative clustering with the constraint to avoid merging clusters associated to different names, is performed. The methods are extended by few biometric models (anchors, some journalists) to directly identify speaker turns and face tracks. The authors validate their approach on the REPERE corpus.

The third paper "*Knowledge Based Query Expansion in Complex Multimedia Event Detection*" (DOI 10.1007/s11042-015-2757-4), co-authored by Maaike de Boer, Klamer Schutte and Wessel Kraaij, compares originally for content-based video retrieval of complex events query expansion methods using common knowledge bases like ConceptNet and Wikipedia to an expert description of the topic. The results show that the expert-based approach is not necessarily better than query expansion using common knowledge, that ConceptNet performs slightly better than Wikipedia and that a late fusion can slightly improve the retrieval performance. The authors conclude that query expansion has a high potential in complex event detection.

The fourth paper "*A Modified Vector of Locally Aggregated Descriptors Approach for Fast Video Classification*" (DOI 10.1007/s11042-015-2819-7), co-authored by Ionut Mironica, Ionut Cosmin Duta, Bogdan Ionescu and Nicu Sebe, investigates a novel perspective for combining frame features for creating a global descriptor. The proposed method combines a modified Vector of Locally Aggregated Descriptors (VLAD) with a Fisher kernel for replacing the classic Bag of Words approach. Additionally, the paper presents a fast algorithm to densely extract global frame features which are easier and faster to compute than classical spatiotemporal local features and a Random Forest approach for replacing the traditional k-means visual vocabulary, allowing a significant speedup. Experiments show the benefit of the proposed methods on different scenarios including: movie genre classification, human action recognition, daily activity recognition, and violence scene classification

The fifth paper "*A scalable summary generation method based on cross-modal consensus clustering and OLAP cube modeling*" (DOI 10.1007/s11042-015-2863-3), co-authored by Gabriel Sargent, Karina R. Perez-Daniel, Andrei Stoian, Jenny Benois-Pineau, Sofian Maabout, Henri Nicolas, Mariko Nakano Miyatakem and Jean Carrive, describes a new scalable video summary generation approach based on On-Line Analytical Processing (OLAP) data cube methods. OLAP methods are brought to the video summary generation by expressing a video within a cross-media feature space and by performing clusterings according to particular subspaces. A large evaluation on a corpus of cultural archives provided by the French Audiovisual National Institute (INA) using information retrieval metrics handling single and multiple reference annotations is performed.

The sixth paper "*Bayes pooling of visual phrases for object retrieval*" (DOI 10.1007/s11042-015-2939-0), co-authored by Wenhui Jiang, hicheng Zhao and Fei Su, addresses the problem of burstiness, i.e., the repetitive occurrence of some certain patterns, in the visual phrases approach for content-based image representation. The authors propose a unified framework for matching geometry-constrained visual phrases and then address the problem of visual phrase burstiness from a probabilistic view by explicitly modelling their distribution for filtering out the bursty ones. The approach is validated in an object retrieval task on five different datasets.

The seventh paper "*Combining Re-Ranking and Rank Aggregation Methods for Image Retrieval*" (DOI 10.1007/s11042-015-3044-0), co-authored by Daniel Carlos Guimarães Pedronette and Ricardo da S. Torres, presents four different approaches for combining re-ranking and rank aggregation methods for Content-Based Image Retrieval. These approaches are evaluated and compared using different visual and textual descriptors and several publicly available image datasets.

The eighth paper "*A Spectrogram-Based Audio Fingerprinting System for Content-Based Copy Detection*" (DOI 10.1007/s11042-015-3081-8), co-authored by Chahid Ouali, Pierre Dumouchel and Vishwa Gupta, presents an audio fingerprinting method for content-based copy detection. The approach is based on the use of multiple thresholds on binarized and pruned spectrograms. Robustness to distortions is achieved using block decomposition and variable sample speed. Experiments show good performance on TRECVid 2009 and 2010 content-based copy detection tasks.

Acknowledgments We would like to thank all authors who submitted their work to this special issue and worked very hard to provide interesting contributions to the field of content-based multimedia indexing. We are also deeply thankful to the many reviewers who did a great job in performing thorough reviews in several review rounds in a timely manner. We hope the readers will enjoy this special issue.

Harald Kosch is a full professor at the University of Passau. He has a computer science engineer diploma (1993) and a PhD in computer science of the École Normale Supérieure de Lyon (1997). From 1997 to 2006, he hold a position of an assistant and later associate professor at the University of Klagenfurt. Since 2006, he leads the Chair of Distributed Information Systems at the University of Passau, where he is responsible for research and teaching in distributed database and information systems and Web databases. His current research activities include multimedia meta-data and databases, multimedia semantics, middleware and Internet applications, linked data and open data.

Georges Quénot is a senior researcher at CNRS (French National Centre for Scientific Research). He has an engineer diploma of the French Polytechnic School (1983) and a PhD in computer science (1988) from the University of Orsay. He currently leads the Multimedia Information Indexing and Retrieval group (MRIM) of the Laboratoire d'Informatique de Grenoble (LIG) where he is also responsible for their activities on video indexing and retrieval. His current research activity includes semantic indexing of image and video documents using supervised learning, networks of classifiers and multimodal fusion.