Multimed Tools Appl (2015) 74:1137–1142
DOI 10.1007/s11042-015-2474-z

GUEST EDITORIAL

Guest editorial: Content-Based Multimedia Indexing

Klaus Schoeffmann · Jenny Benois-Pineau · Bernard Merialdo · Tamás Szirányi

Published online: 30 January 2015

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, image analysis and human computer interaction have contributed to the success of multimedia systems. Although, there has been significant progress in the field, we still face situations when the system show 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.

We received 48 submissions, but only accepted 17 (35.42 %), according to the review process that was very rigorous for this special issue and consisted of several rounds of review. The high number of submissions reflect the importance and timeliness of the research field on content-based multimedia indexing. The selected 17 papers in this special issue are high-class publications and cover a wide range of problems in content indexing of multimedia data.

The first paper "*Variability modelling for audio events detection in movies*" (DOI 10.1007/s11042-014-2038-7), co-authored by Cédric Penet, Claire-Hélène Demarty, Guillaume Gravier, and Patrick Gros, proposes to model the variability between the soundtracks of Hollywood movies using a factor analysis technique, which is then used to compensate the audio features. For that purpose they use multiple audio words sequences and contextual Bayesian networks.

K. Schoeffmann
Institute of Information Technology, Klagenfurt University, Klagenfurt, Austria
e-mail: ks@itec.aau.at

J. Benois-Pineau (✉)
LaBRI, University of Bordeaux, Bordeaux, France
e-mail: benois-p@labri.fr

B. Merialdo
EURECOM, Multimedia Communications, Sophia Antipolis, France
e-mail: Bernard.Merialdo@eurecom.fr

T. Szirányi
Distributed Events Analysis Research Laboratory, Institute for Computer Science and Control (MTA SZTAKI), Budapest, Hungary
e-mail: sziranyi@sztaki.hu

The second paper is entitled "*Combining content with user preferences for non-fiction multimedia recommendation: a study on TED lectures*" (DOI 10.1007/s11042-013-1840-y), written by Nikolaos Pappas and Andrei Popescu-Belis. It provides a comparison of keyword-based (TFIDF) and semantic vector space methods (LSI, LDA, RP, and ESA) for personal recommendation of videos from the TED dataset.

The third paper "*Large scale classifiers for visual classification tasks*" (DOI 10.1007/s11042-014-2049-4), co-authored by Thanh-Nghi Doan, Thanh-Nghi Do, and Francois Poulet, addresses the problem of training fast and accurate visual classifiers on several multi-core computers. They evaluate their method with the 100 largest classes of ImageNet and ILSVRC 2010 and show that their approach, which extends state-of-the-art linear (LIBLINEAR-CDBLOCK) and non-linear classifiers (Power Mean SVM), can save up to 82.01 % memory consumption and is much faster than the original implementation.

The fourth paper is co-authored by Abdelkader Hamadi, Philippe Mulhem, and Georges Quénot, and about "*Extended conceptual feedback for semantic multimedia indexing*" (DOI 10.1007/s11042-014-1937-y). This paper addresses the problem of detecting a large number of visual concepts in images or video shots. The proposed "conceptual feedback" considers the relations between concepts in order to improve the overall concept detection performance. The authors further propose three extensions of their method and evaluate them in context of the TRECVID 2012 SIN (semantic indexing) task.

The fifth paper is entitled "*Classification of Alzheimer's disease subjects from MRI using hippocampal visual features*" (DOI 10.1007/s11042-014-2123-y), co-authored by Olfa Ben Ahmed, Jenny Benois-Pineau, Michèle Allard, Chokri Ben Amar, and Gwénaelle Catheline. Their work is about using visual features from the most involved region (hippocampal area) in Alzheimer's Disease (AD). They propose a late fusion method to increase precision results (85 % - 87 % accuracy). Magnetic Resonance images taken from 218 subjects were used for their evaluation.

The sixth paper, written by Bahjat Safadi, Nadia Derbas, and Georges Quénot, is about "*Descriptor optimization for multimedia indexing and retrieval*" (DOI 10.1007/s11042-014-2071-6). This work proposes to combine a PCA-based dimensionality reduction method with pre- and post-PCA non-linear transformations, in order to allow for usage in large-scale systems.

The next paper, "*Best practices for learning video concept detectors from social media examples*" (DOI 10.1007/s11042-014-2056-5), co-authored by Svetlana Kordumova, Xirong Li, and Cees G. M. Snoek. The authors investigate how to learn video concept detectors from social media sources, such as Flickr and YouTube. For that purpose they investigate three strategies for positive example selection, three strategies for negative example selection, and three learning strategies, and evaluate these methods with the TRECVID 2012 dataset.

Anna Llagostera Casanovas and Andrea Cavallaro present their work on "*Audio-visual events for multi-camera synchronization*" (DOI 10.1007/s11042-014-1872-y). They propose a multimodal method for automatic synchronization of audio-visual recordings captured from independent cameras, which jointly processes data from audio and video channels to estimate inter-camera delays that are used to temporally align the recordings. Their results show that they can outperform other methods working with audio-only or video-only approaches.

The next paper is about "*Learning latent semantic model with visual consistency for image analysis*" (DOI 10.1007/s11042-014-1916-3), co-authored by Jian Cheng, Peng Li, Ting Rui, and Hangqing Lu. In this work the authors propose to use both the topic consistency and word consistency in semantic space to adapt the traditional PLSA model to the visual content analysis task.

Ricardo C. Sperandio, Zenilton K. G. Patrocínio Jr., Hugo B. de Paula, and Silvio J. F. Guimaraes present their work about "*An efficient access method for multimodal video*

retrieval" (DOI 10.1007/s11042-014-1917-2). They propose the Slim^2-tree, which is an effective and efficient content-based video retrieval technique that allows using multiple modalities within a single index structure. It is capable of using different distance measures and can perform both multimodal and unimodal search.

The next paper is entitled "*Lexical speaker identification in TV shows*" (DOI 10.1007/s11042-014-1940-3) and written by Anindya Roy, Hervé Bredin, William Hartmann, Viet Bac Le, Claude Barras, and Jean-Luc Gauvain. It investigates the problem of speaker identification in recordings of conversations, debates, discussions, and Q & A sessions (REPERE corpus) by lexical information extraction. More precisely, they study four lexical speaker identification approaches in this paper, including TFIDF, BM25, and LDA-based topic modeling.

"*A generic framework for semantic video indexing based on visual concepts/contexts detection*" (DOI 10.1007/s11042-014-1955-9) is presented by Nizar Elleuch, Anis Ben Ammar, and Adel M. Alimi. The authors present a video indexing scheme consisting of three levels: (1) low-level processing, such as shot boundary detection, (2) semantic models for supervised learning of concepts/contexts, and (3) semantic interpretation of concepts/contexts by exploiting fuzzy knowledge.

Hong-Mei Hou, Xin-Shun Xu, Gang Wang, and Xiao-Lin Wang present their work about "*Joint-Rerank: a novel method for image search reranking*" (DOI 10.1007/s11042-014-1962-x). Their proposed image reranking framework considers multiple modalities of images through a multigraph, where each image is a node with multimodal attributes (textual and visual cues) and the edges between nodes express both intra-modal and inter-modal similarities of images.

The next work is about "*Data-driven approaches for social image and video tagging*" (DOI 10.1007/s11042-014-1976-4), and co-authored by Lamberto Ballan, Marco Bertini, Tiberio Uricchio, and Alberto Del Bimbo. In this paper the authors review state-of-the-art approaches to automatic annotation and tag refinement for social images, which address the problem of how to deal with the low quality of the available metadata.

Cong Bai, Jinglin Zhang, Zhi Liu, and Wan-Lei Zhao present a work about "*K-Means based histogram using multiresolution feature vectors for color texture database retrieval*" (DOI 10.1007/s11042-014-2053-8). More precisely, they propose a k-means based histogram (KBH) using a combination of color and texture features for the field of image retrieval. Their approach uses multiresolution feature vectors generated from coefficients of Discrete Wavelet Transform (DWT). The vector space is partitioned with K-means and finally the KBHs are fused using z-score normalized Chi-Square distance.

The next paper is entitled "*Content-based Singer Classification on Compressed Domain Audio Data*" (DOI 10.1007/s11042-014-2189-6) and co-authored by Han Tsung Tsai, Siang Yu Huang, Pei-Yun Liu, and Ming De Chen. More specifically, a singer identification approach to automatically identify the singer of an unknown MP3 audio data is proposed in this paper. The approach works in MP3 compressed domain using Mel-Frequency Cepstral Coefficients (MFCC) as feature and Gaussian mixture model (GMM) for describing the distribution of the MFCC vector.

The last paper in this special issue is about "*3D model retrieval based on linear prediction coding in cylindrical and spherical projections using SVM-OSS*" (DOI 10.1007/s11042-014-2055-6) and co-authored by Vahid Mehrdad and Hossein Ebrahimnezhad. They present a 3D model descriptor based on linear prediction coding (LPC) coefficients to retrieve 3D objects. To improve retrieval performance they employ an SVM-OSS similarity measure to efficiently compare two feature vectors. In the evaluation the method is compared to other current methods.

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. Moreover, we are 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.

Klaus Schoeffmann is assistant professor at the Institute of Information Technology at Alpen-Adria-Universität Klagenfurt, Austria, where he received his Ph.D. degree in 2009. His current research focuses on visual content analysis and interactive video search. He is the author of numerous peer-reviewed publications on video browsing, video exploration, and video content processing. Klaus Schoeffmann has co-organized international conferences, special sessions and workshops (e.g., MMM 2012, CBMI 2013). He is co-founder of the Video Search Showcase (VSS), formerly known as Video Browser Showdown (VBS). He is member of the IEEE and the ACM and a frequent reviewer for international conferences and journals in the field of Multimedia.

Jenny Benois-Pineau is a full professor of Computer science at the University Bordeaux and chair of Video Analysis and Indexing research group in Image and Sound Department of LABRI UMR 58000 Université Bordeaux/CNRS/ENSEIRB. She is also a deputy scientific director of theme B of French national research unity GDR CNRS ISIS. She obtained her PhD degree in Signals and Systems in Moscou and her Habilitation à Diriger la Recherche in Computer Science and Image Processing from University of Nantes France. Her topics of interest include image and video analysis and indexing, motion analysis and content description for content-based multimedia retrieval. She is the author and co-author of more than 110 papers in international journals, conference proceedings, book chapters. She has tutored an co-tutored 20 PhD students. She is associated editor of EURASIP Signal Processing: Image Communication, Elsevier, Multimedia Tools and applications, Springer. She has served in numerous program committees in international conferences and workshops: ACM MM, CIVR, CBMI, AMR, IPTA, SAMT, ECMCS. She has served as expert for European Commission since FP4 and was a

member of Technical Advisory group for Media programme EACEA DG Culture, CE. She is a member of Multimedia Commission of French Ministry of National Education and member of scientific board of International Center for Mathematical Modelling at the University of Växjo, Sweden. She has been coordinator or leading researcher in EU funded, bi-lateral and national research projects.

Bernard Merialdo is professor in the Multimedia Department of EURECOM, France and current head of the department. A former student of the Ecole Normale Supérieure, Paris, he received a PhD from Paris 6 University and a "Habilitation à Diriger des Recherches" from Paris 7 University. For more than 10 years, he was a research staff, then project manager at the IBM France Scientific Center, working on probabilistic techniques for Large Vocabulary Speech Recognition. He later joined EURECOM to set up the Multimedia Department. His research interests are the analysis, processing, indexing and filtering of Multimedia information to solve user-related tasks. His research covers a whole range of problems, from content extraction based on recognition techniques, content understanding based on parsing, multimedia content description languages (MPEG7), similarity computation for applications such as information retrieval, user personalization and user interaction for the design of innovative applications. He participates in numerous conference program committees. He is part of the organizing committee for the CBMI workshop series. He was editor for the IEEE Transactions on Multimedia and general chair of the ACM Multimedia conference in 2002. He often acts as an expert and reviewer for French and European research programs. He is a Senior Member of IEEE and member of ACM.

Tamás Szirányi leads the Distributed Events Analysis Research Laboratory at the Institute for Computer Science and Control, Hungarian Academy of Sciences. His research activities include machine perception, pattern recognition, texture and motion segmentation, Markov Random Fields and stochastic optimization, remote

sensing, surveillance, intelligent networked sensor systems, graph based clustering, digital film restoration. Dr. Szirányi was the founder and past president (1997 to 2002) of the Hungarian Image Processing and Pattern Recognition Society. He was an Associate Editor of IEEE T. Image Processing (2003–2009), and he has been an AE of Digital Signal Processing since 2012. He was honored by the Master Professor award in 2001, by the Széchenyi professorship and the ProScientia (Veszprem) award in 2011. He is a Fellow both of the IAPR and the Hungarian Academy of Engineering from 2008. He has more than 240 publications including 45 in major scientific journals.