Multimed Tools Appl (2011) 55:1–6
DOI 10.1007/s11042-010-0618-8

GUEST EDITORIAL

Introduction to the special issue on image and video retrieval: theory and applications

Ioannis Kompatsiaris · Stephane Marchand-Maillet · Roelof van Zwol · Sébastien Marcel

Published online: 7 October 2010

This editorial introduces the Special Issue, which contains a selection of papers that present recent advances in a number of areas relating to image and video annotation and retrieval. Image and Video retrieval have now reached a state where successful techniques and applications have begun flourishing and technologies for retrieving multimedia content enable new applications and services for both commercial and personal end users. While existing techniques already provide promising results, recent advances such as the availability of large amounts of user-related data pose new challenges and generate the need for novel approaches to efficiently exploit them. Such data can be generated by the users explicitly, for example through tagging and participation in social media sharing properties, or implicitly, by analyzing user logs and feedback. In addition, affordable sensors provide further information, e.g. GPS data that can be used to increase the efficiency of existing solutions, such as concept detection. The aim of this special issue is to provide an overview of current research in emerging topics of image and video retrieval models and applications, while at the same time, including papers advancing the state of the art in related mature issues.

I. Kompatsiaris (✉)
Informatics and Telematics Institute, Centre for Research and Technology Hellas,
Thermi-Thessaloniki, Greece
e-mail: ikom@iti.gr

S. Marchand-Maillet
University of Geneva, Geneva, Switzerland
e-mail: Stephane.Marchand-Maillet@unige.ch

R. van Zwol
Yahoo! Research, Barcelona, Spain
e-mail: roelof@yahoo-inc.com

S. Marcel
Idiap Research Institute, Martigny, Valias, Switzerland
e-mail: Sebastien.Marcel@idiap.ch

This Special Issue attempts to present a representative sample of ongoing research focusing mainly on multimodal approaches that combine automatic analysis and user feedback, and that exploit explicit and implicit user activity for machine learning and annotation, face recognition and interactive applications for retrieval and surveillance. The Special Issue Call for Papers received a strong response from the community. A total of 22 high quality manuscripts were submitted for consideration. Of these submissions, 7 papers were accepted following a rigorous review process, coordinated by the guest editors. We are grateful to all authors for their interest and to the reviewers for their effort to ensure the highest possible quality in all accepted papers.

Tagging-Annotation The problem of obtaining accurate access to digital media is not only a challenging but also critical problem in many situations. In many approaches automatic image annotation is achieved by using supervised learning, which is performed by concept classifiers that have been trained on labelled example images. Such approaches usually involve an expensive manual annotation effort for the generation of concept training data. In several papers of the special issue, alternative approaches are examined for image and video annotation, exploiting explicit and implicit user feedback. In "*Methods for automatic and assisted image annotation*", Rui Jesus, Arnaldo Abrantes and Nuno Correia propose methods for facilitating the annotation of digital memories such as personal photo collections. The process starts with an automated annotation strategy making use of all possible modes available (including audio tags and GPS data) to detect relevant concepts in the digital content. The authors then propose the use of a social game to help the system enhance its performance, while replacing the tedious task of image-keyword association with an enjoyable gaming action. Scores computed from the efficacy and reliability of the user actions and also a history of the system usage encourage users to provide semantic feedback. In turn, this feedback is used to enhance models within the automated annotation part, initiating a virtuous cycle, that finally leads to an increased level of semantics within the collection. Authors validate their approach with user studies assessing the usefulness and usability of their solutions. Such a study on means to augment the semantic description of multimedia content is a relevant complement to proposals over search-based systems.

In "*Reliability and effectiveness of clickthrough data for automatic image annotation*", Theodora Tsikrika, Christos Diou, Arjen P de Vries and Anastasios Delopoulos propose the use of clickthrough data collected from search logs as a source for the automatic generation of concept training data, thus avoiding the expensive manual annotation effort. They investigate and evaluate this approach using a collection of 97,628 photographic images. The results indicate that the contribution of search log based training data is positive despite their inherent noise; in particular, the combination of manual and automatically generated training data outperforms the use of manual data alone. It is therefore possible to use clickthrough data to perform large-scale image annotation with little manual annotation effort or, depending on performance, using only the automatically generated training data.

Zhineng Chen, Juan Cao, Tian Xia, Yicheng Song, Yongdong Zhang and Jintao Li in their paper titled "*Web Video Retagging*" propose a method for retagging video to address the problem of incomplete, imprecise and unranked tagging. Compared to state-of-the-art, this method handles all three problems simultaneously, whereas competitive approaches deal with one or two of the aforementioned issues. In the proposed video retagging, neighbor tags are first collected and considered as possible relevant tags. Their tag

relevance is simultaneously estimated from both global and individual perspectives. In particular, given a web video, video retagging first collects its textually and visually related neighbors. All tags attached to the neighbors are treated as possibly relevant, after which a ranked retagged tag list is generated by inferring the degree of relevance of these tags from both global and video-specific perspectives, using two different graph based models. The authors evaluate their approach from both application-oriented and user-based perspectives. They conclude that in most cases, the ranked retagged tag list is better than original set of tags, in terms of completeness, precision and ranking.

Face recognition Nowadays, research in face recognition is mature enough for many applications but mainly for multimedia indexing and retrieval tasks. Currently, several companies have deployed this technology into products such as Apple iPhoto and Google Picasa. However, in those systems it is not clear if it is the user feedback or the face recognition technology that has an influence on the retrieval performance. Moreover, it is observed that the performance of face recognition itself could still be improved as it is a process affected by face pose, illumination and occlusion.

The two papers, namely "Person re-identification in TV series using robust face recognition and user feedback" by Mika Fischer, Hazım K Ekenel and Rainer Stiefelhagen and "Spatio-Temporal Tube data representation and Kernel design for SVM-based video object retrieval system" by Shuji Zhao, Frédéric Precioso and Matthieu Cord address respectively the two above issues. The first one proposes an alternative method to face retrieval in media collections and more specifically in videos. It is shown to be robust to pose, illumination and conditions. The second paper describes a novel spatio-temporal representation and classification method for face recognition in videos.

Interactive applications It is not always possible for automated video analysis to produce reliable results and detect unknown events, which can be used in real applications. One practical approach is to watch all the recorded video data in fast-forward mode, if possible. The paper "*Information-Based Adaptive Fast-Forward for Visual Surveillance*" by Benjamin Höferlin, Markus Höferlin, Daniel Weiskopf and Gunther Heidemann presents a method to adapt the playback velocity of the video to the temporal information density, so that the users can explore the video under controlled cognitive load. The proposed approach can cope with static changes and is robust to video noise. They first formulate temporal information as symmetrized Rényi divergence, deriving this measure from signal coding theory. Further, they discuss the animated visualization of accelerated video sequences and propose a physiologically motivated blending approach to cope with arbitrary playback velocities. Their experimental results including both objective and user-based subjective evaluation and comparisons show the advantages over motion-based measures.

Interactive video retrieval has also attracted a lot of interest resulting into significant advances in content-based indexing, concept detection, multimodal approaches, interfaces and visualization. Understanding user needs is an important requirement for efficient search and therefore many approaches make use of explicit and implicit user feedback, with implicit feedback strategies gaining more attention recently. In the paper "*Towards Hierarchical Context: Unfolding Visual Community Potential for Interactive Video Retrieval*", Lin Pang, Juan Cao, Lei Bao, Yongdong Zhang and Shouxun Lin, propose a hierarchical community-based feedback algorithm. They exploit the visual community structure in a visual-temporal correlation network and utilize it to improve interactive video

retrieval. By re-ranking the video shots through diffusion processes respectively at the inter-community and intra-community level, the feedback algorithm can make full use of the limited user feedback. Furthermore, since it avoids the computation of the entire graph, the feedback algorithm can make quick responses to user feedback, which is particularly important for large video collections. In complement, the authors propose a community-based visualization interface called *VideoMap*. By organizing the video shots following the community structure, the VideoMap presents a comprehensive and informative view of the whole dataset to facilitate users' access. Moreover, the VideoMap can help users to quickly locate the potential relevant regions and make active annotations according to the distribution of labeled samples on the VideoMap.

By its rich content, this Special Issue demonstrates the broad diversity and the lively aspect of Image and Video Retrieval research. We believe that it includes a selection of papers that present recent representative exciting results in the field. From these papers it is clear that there is a significant focus of the community to consider user activities, feedback and interaction of any form in their approaches. Issues such as fusion and multimodality, handling of large-scale noisy data, modeling user actions and expectations need to be addressed, which still offers many opportunities for research. A more recent trend is the inclusion of community-based models in the analysis and management of distributed media. These so called social media related applications and data play an important role in the field, either as rich sources of content and user activities or as new means to generate data (e.g. social games). We see this as an important step in order to meet the demanding user-centric challenges of future multimedia services and applications. As touched upon in contributions from this Special Issue, questions of scalability in indexing and retrieval or of higher-level semantic access to digital content cannot find a solution without considering user community models as part of their responses.

Ioannis (Yiannis) Kompatsiaris is a Senior Researcher (Researcher B') with the Informatics and Telematics Institute. His research interests include semantic multimedia analysis, indexing and retrieval, Web 2.0 content analysis, knowledge structures, reasoning and personalization for multimedia applications. He received his Ph.D. degree in 3-D model based image sequence coding from the Aristotle University of Thessaloniki in 2001. He is the coauthor of 40 papers in refereed journals, 20 book chapters, 4 patents and more than 150 papers in international conferences. He has served as a regular reviewer for a number of international journals and conferences and he has been the co-organizer of various conferences and workshops, such as the ACM CIVR, WIAMIS and SSMS. He is the coordinator of the WeKnowIt—Emerging, Collective Intelligence for personal, organisational and social use European Integrated Project. Recently, he has been appointed as Chair of the Technical Committee 14 of the International Association for Pattern Recognition (IAPR-TC14, "Signal Analysis for Machine Intelligence"). He is a member of IEEE and ACM.

Stéphane Marchand-Maillet has founded and is heading the Viper group (http://viper.unige.ch) in the Department of Computer Science at University of Geneva. His research is directed towards multimedia information retrieval with emphasis on Multimedia Content Abstraction, *ie* attaching semantic information to multimedia documents at cheapest cost. In particular, he is interested in all aspects related to multimedia information mining and retrieval and smooth acquisition of knowledge by enhancing user or group interaction. He has been appointed as Chair of the Technical Committee 12 of the International Association for Pattern Recognition (IAPR-TC12, "Multimedia and Visual Information Systems", http://www.iapr-tc12.org). Recently, he was the general co-chair of the International Conference of the ACM-SIG on Information Retrieval in 2010 (ACM-SIGIR 2010, http://www.sigir2010.org). He has served as a regular reviewer for a number of international journals and conferences. He is a member of the ACM (SIGIR and SIGMM).

Dr. Roelof van Zwol is a senior research scientist, and leads the Multimedia group at Yahoo! Research since he joined Yahoo! in August 2006. His research focuses on media interaction, mining and search, and delivering innovation in Yahoo!'s multimedia products. He has more than 60 peer-reviewed publications, many of which appeared on premier conferences such as ACM Multimedia, WWW, and SIGIR. He has been the co-organizer of various conferences and workshops, such as ACM CIVR, the Multimedia Information Retrieval workshop at SIGIR 2007, Future of Web Search workshop, and Dutch/Belgian Information Retrieval workshop. He was the technical coordinator of the SEMEDIA European Project on Search Environments for Multimedia (FP6-045032), where the objective was is to develop models for media retrieval by incorporating ideas of social relevance. He currently manages Yahoo!'s involvement in the WeKnowIt EU project. Prior to joining Yahoo!, he was an assistant professor at Utrecht University, working on spatial information retrieval and XML retrieval. At that time he also participated in the SPIRIT EU project on geographic information retrieval, and has lead the Multimedia track of INEX (Initiative for the Evaluation of XML Retrieval). He obtained his PhD. degree from the Database group at University of Twente, the Netherlands on the topic of "Modelling and Searching Web-based Document Collections."

Sébastien Marcel is senior research scientist at the Idiap Research Institute. He is interested in multi-modal biometric person recognition, man–machine interaction and content-based multimedia indexing and retrieval. He has obtained his PhD in signal processing from "Université de Rennes I" in France (2000) at CNET, the research center of France Telecom (now Orange Labs). He currently leads the Biometric Person Recognition research team at the Idiap Research Institute and manages collaborative (CH and EU FP7) research projects. In 2010, he was appointed Visiting Professor at the University of Cagliari (IT) where he taught a series of lectures on "face recognition". He has served as a regular reviewer for a number of International journal and conferences. He is also a member of IEEE.