GUEST EDITORIAL

Trends in semantic and digital media technologies

**Marcin Grzegorzek · Michael Granitzer ·
Stefan Rüger · Michael Sintek · Thierry Declerck ·
Massimo Romanelli**

The objective of this special issue is to report on recent trends in digital and media technologies responding to the challenges of managing and accessing multimedia (images, audio, video, 3D/4D material, etc.). In a highly selective review procedure we accepted contributions describing recent work that aims at narrowing the large disparity between the low-level multimedia descriptors and the richness of subjectivity of semantics in user queries and human interpretations of audiovisual data.

The articles in this special issue can be grouped into following categories: Multimedia Analysis (Section 1), Multimedia Ontologies and Data Integration (Section 2), and Social Media and Retrieval (Section 3).

M. Grzegorzek (✉)
Institute for Vision and Graphics, University of Siegen, Siegen, Germany
e-mail: marcin.grzegorzek@uni-siegen.de

M. Granitzer
Institute for Media Informatics, University of Passau, Passau, Germany
e-mail: michael.granitzer@uni-passau.de

S. Rüger
Knowledge Media Institute, The Open University, Milton Keynes, UK
e-mail: s.rueger@open.ac.uk

M. Sintek
Knowledge Management Department, DFKI GmbH, Kaiserslautern & Berlin,
Berlin, Germany
e-mail: michael.sintek@dfki.de

T. Declerck
Language Technology Lab, DFKI GmbH, Saarbrücken, Saarbrücken, Germany
e-mail: declerck@dfki.de

M. Romanelli
Intelligent User Interfaces Department, DFKI GmbH, Saarbrücken, Saarbrücken,
Germany
e-mail: massimo.romanelli@dfki.de

1 Multimedia analysis

The predominant mode of operation in multimedia analysis has traditionally been bottom-up, i.e., working out high-level semantic features from low-level and easy to compute characteristics of the multimedia representation. In this special issue only one of three papers in the area of multimedia analysis follows this path (emotion prediction [1], described below). The second paper in multimedia analysis uses a high-level desired representation of a film script and maps backwards and forwards to image-processing output of film scenes [2]. Finally, the last paper in this section uses high-level semantics to guide the low-level task of keyframe detection [9] in a reversal of traditional workflow.

Ricardo Calix and Gerald Knapp propose a methodology for actor level emotion magnitude prediction in text and speech [1]. Their model feeds the previous predicted emotion magnitudes into the model for the current state that is also influenced by particular current text and speech features. Their methodology is drawn from machine learning, i.e., selecting and fitting a model to the available data. They find that nonlinear Support Vector Regression models with radial basis functions work best and identify the contribution of the selected features to the prediction.

In an example of high-level conceptual modelling that is brought to bear on low-level features, Benjamin Diemert, Ana Pinzari, Claude Moulin, Marie-Helene Abel and Marcus Shawky [2] define conceptual models of film scripts that represent information about realising the film shooting for a) prior guidance of the (amateur) cameraman b) feedback during shooting based on image-processing algorithms and c) after the shooting for selection of appropriate video sequences. Diemert et al.'s work on mapping semantic scripts to the output of image processing algorithms ultimately aims at leveraging amateur video material in a professional production.

Suet-Peng Yong, Jeremiah Deng and Martin Purvis hypothesise that visual salience can be better modelled using top-down mechanisms that incorporate object semantics than without [9]. They predict that the approach of automated semantics extraction can be used to improve video summarisation, indexing and retrieval. In their own work they have developed a framework that models semantic contexts for keyframe extraction and monitors sequential changes in the semantic context. This results in an ability to detect and locate significant novelty in the video stream. In their concrete application, Yong et al. first segmented wildlife video frames, extracted features and matched image blocks in order to automatically construct a co-occurrence matrix of labels that ultimately represent the semantic context of a scene. They demonstrated that using this approach yields better keyframe extraction than using only the low-level features.

2 Multimedia ontologies and data integration

Ontologies play an important role in multimedia by allowing to exchange the semantics of multimedia content between distributed information systems. In this special issue three papers discuss the role of ontologies and the integration of distributed multimedia sources. The authors of [7] compare the landscape of multimedia ontologies in the last decade, while [8] proposes and compares two ontology matching techniques to semantic image retrieval. In contrast, the authors of [3] present an mediator based approach to integrate multiple heterogeneous data sources.

Mari Carmen Suarez-Figueroa, Ghislain Auguste Atemezing, and Oscar Corcho compare well-known ontologies in the multimedia domain [7]. Therefore, they introduce a framework called FRAMECOM-MON, which takes process-oriented dimension, such as the methodological one, outcome-oriented dimensions, like multimedia aspects, understandability, and evaluation criteria into account.

Konstantin Todorov, Nicolas James, and Celine Hudelot present an approach for multimedia ontology matching with the aim to associate common-sense knowledge to multimedia concepts [8]. Therefore, the authors extend textual concept matching approaches by visual representations of images and introduce a new multi-modal graph matching method. Their evaluation yields the conclusion that textual and visual modalities have to be seen as complementary to each other.

Claudio Gennaro, Domenico Beneventano, Sonia Bergamaschi, and Fausto Rabitti describe a mediator-based approach for providing a unified multimedia access service over integrated heterogeneous multimedia sources [3]. Their approach builds on the MOMIS integration system and the MILOS multimedia data management system which facilitate query-by-example access to multimedia collections. The argued strong points of this approach are MILOS flexible similarity measures and the MOMIS mediators exploitation of ranks of local answers.

3 Social media and retrieval

Remaining group of papers in this special issue concentrate on social media and retrieval. While in [5] an image tag ranking system called i-TagRanker is presented, [4] describes an approach for image retrieval using spatial relations. Further, [6] focuses on multiple perspective interactive search defining a paradigm for exploratory search and information retrieval on the web.

Jin-Woo Jeong, Hyun-Ki Hong, and Dong-Ho Lee propose in [5] a novel image tag ranking system called i-TagRanker which exploits the semantic relationships between tags for re-ordering the tags according to the relevance with an image. The proposed system works of two phases, namely the tag propagation phase, and the tag ranking phase. In the tag propagation phase, the most relevant tags from similar images are collected and propagated to an untagged image. In the tag ranking phase, tags are ranked according to their semantic relevance to the image. The system has been evaluated on a Flickr photo collection with over 30,000 images.

Carlos Arturo Hernandez-Gracidas, Luis E. Sucarand, and Manuel Montes-y-Gomez report in [4] on their work on improving image retrieval by using spatial relations. They model the spatial relations with conceptual graphs. Additionally, they propose a term-weighting scheme and use more than one sample image for retrieval applying several late fusion techniques. Their methods were evaluated with a rich and complex image dataset, based on the 39 topics developed for the ImageCLEF 2008 photo retrieval task.

Rahul Singh, Ya-Wen Hsu, and Naureen Moon focus in their article [6] on multiple perspective interactive search defining a paradigm for exploratory search and information retrieval on the web. Their system allows simultaneous and semantically correlated presentation of query results from different semantic perspectives. Users can explore the results either using a specific perspective or through a combination of perspectives via highly-intuitive yet powerful interaction operators. In the proposed paradigm, hits obtained from executing a query are first analysed to determine latent

content-based correlations between the pages. Next, the pages are analysed to extract different types of perceptual and informational cues. This information is used to organise and present the results through an interactive and reflective user interface which supports both exploration and search.

References

1. Calix R, Knapp GM (2011) Actor level emotion magnitude prediction in text and speech. Multimed Tools Appl. doi:10.1007/s11042-011-0909-8
2. Diemert B, Pinzari A, Moulin C, Abel MH, Shawky MM (2011) Mapping semantic script with image processing algorithms to leverage amateur video material in professional production. Multimed Tools Appl. doi:10.1007/s11042-011-0908-9
3. Beneventano D, Gennaro C, Bergamaschi S, Rabitti F (2011) A mediator-based approach for integrating heterogeneous multimedia sources. Multimed Tools Appl. doi:10.1007/s11042-011-0904-0
4. Hernandez-Gracidas CA, Sucar LE, y Gomez MM (2011) Improving image retrieval by using spatial relations. Multimed Tools Appl. doi:10.1007/s11042-011-0911-1
5. Jeong JW, Hong HK, Lee DH (2011) i-TagRanker: an efficient tag ranking system for image sharing and retrieval using the semantic relationships between tags. Multimed Tools Appl. doi:10.1007/s11042-011-0903-1
6. Singh R, Hsu YW, Moon N (2011) Multiple perspective interactive search: a paradigm for exploratory search and information retrieval on the web. Multimed Tools Appl. doi:10.1007/s11042-011-0910-2
7. Suarez-Figueroa MC, Atemezing GA, Corcho O (2011) The landscape of multimedia ontologies in the last decade. Multimed Tools Appl. doi:10.1007/s11042-011-0905-z
8. Todorov K, James N, Hudelot C (2011) Multimedia ontology matching by using visual and textual modalities. Multimed Tools Appl. doi:10.1007/s11042-011-0912-0
9. Yong SP, Deng JD, Purvis MK (2011) Wildlife video key-frame extraction based on novelty detection in semantic context. Multimed Tools Appl. doi:10.1007/s11042-011-0902-2

Marcin Grzegorzek　is assistant professor at the University of Siegen where he leads the Research Group for Pattern Recognition. Currently, he is also visiting professor at the University of Economics in Katowice. He received his PhD with distinction from the University Erlangen-Nuremberg in

2007. From 2006 to 2008 he was research assistant at the Queen Mary, University of London. From 2008 to 2010 he worked as lecturer for the University Koblenz-Landau. Marcin was general chair of the SAMT 2010 conference, the 5th International Conference on Semantic and Digital Media Technologies. Moreover, he acts as secretary in the Executive Board of the SMaRT (Semantic Multimedia Research and Technology) Association and is associate of the Think Thank "Stiftung Neue Verantwortung". His research interests focus on object recognition, multimedia retrieval, knowledge-based image processing, and behavioural biometry.

Michael Granitzer is professor for Media Informatics at the University of Passau, Germany. Before he was Scientific Director of the Know- Center Graz and assistant professor at the Knowledge Management Institute of Graz University of Technology, where he worked on topics like automatic text classification, large scale text retrieval and information visualisation. As a Marie Curie Research Fellow at Mendeley Ltd., he investigated machine learning and information retrieval techniques for managing large academic knowledge bases in 2011. His research addresses topics in the field of Information Retrieval, Machine Learning, Visual Analytics, Knowledge Discovery, Semantic Technologies and Multimedia Semantics. Granitzer studied Telematik at Graz University of Technology with special focus on Computational Intelligence and holds a PhD degree, passed with distinction, in technical science. He published around 110 mostly peer-reviewed publications including journal publications, book chapters and books in the above mentioned fields. He was member in several program committees and reviewer for well known journals like Elseviers Pattern Recognition, Springers Journal on Multimedia Tools and Applications and the IEEE Journal on Transactions of Knowledge and Data Engineering. He has been involved in several nationally and internationally funded research projects. In particular, he led work packages in the nationally funded FIT-IT Project MISTRAL (Measurable, Intelligent and reliable Semantic exTraction and RetrievAL of multimedia data), managed the 4 times awarded FIT-IT project DYONIPOS (DYnamic ONtology based Integrated Process OptimiSation), has been scientific coordinator of the Marie Curie FP 7 IAPP Project TEAM (Transferring knowledgE in Academic knowledge) 4 and of the Marie Curie FP 7 IRSES project WIQ-EI (Web Information Quality Evaluation Initiative) and is principal investigator of the FP 7 STREP CODE Commercially empowered Linked Open Data Ecosystems in Research.

Stefan Rüger Physics at Freie Universität Berlin and gained his PhD at Technische Universität Berlin (1996). He carved out his academic career at Imperial College London (1997–2006), where he also held an EPSRC Advanced Research Fellowship (1999–2004). In 2006 he became a professor of Knowledge Media when he joined The Open University's Knowledge Media Institute to cover the area of Multimedia and Information Systems. Since 2009 he has held a Honorary Professorship from the University of Waikato, New Zealand, for his collaboration with the Greenstone Digital Library group on Multimedia Digital Libraries.

Rüger's research interests are firmly rooted in the area of automated multimedia understanding with a view to improving multimedia indexing, search and browsing in digital collections and beyond. The relevant research contributes to—and is informed by—machine learning, computer vision and artificial intelligence. Amongst other projects, he was principal investigator in the EPSRC-funded Multimedia Knowledge Management Network, of an EPSRC grant to research and develop video digital libraries and principal investigator for The Open University in the European FP6-ICT project PHAROS that established a horizontal layer of technologies for large-scale audio-visual search engines. In the past decade he has served the academic community in various roles as conference chair, programme chair, posters chair, journal editor, guest editor and as referee for well over 100 organisations, be Computing journals, international conferences and research sponsors. Rüger is a member of the EPSRC College, ACM, BCS, the BCS IRSG committee and a fellow of the Higher Education Academy.
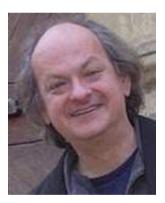

Michael Sintek (http://www.dfki.uni-kl.de/~sintek/) studied computer science and economics at the University of Kaiserslautern and received the Diplom (master's degree) in 1996. Since then, he is working as a research scientist at the German Research Center for Artificial Intelligence (DFKI

GmbH). As a visiting researcher at the Stanford Medical Informatics department and the Stanford Database Group, he developed various Semantic Web extensions for the the frame-knowledge acquisition tool Protege and co-developed the Semantic Web rule language TRIPLE. Recently, he worked on semantic technologies in the BMWI-funded lighthouse project THESEUS.

Currently, he is, together with Thierry Declerck, co-head of the Competence Center Semantic Web (CCSW, http://semanticweb.dfki.de/) at DFKI.

His main scientific interests are knowledge representation/ontologies and reasoning (esp. rule-based) in the context of Semantic Web and Linked Data.

Thierry Declerck is a senior consultant at he Language Technology Lab of the German Research Centre for Artificial Intelligence (DFKI GmbH). At DFKI he worked first in the field of Information Extraction. He was later the responsible scientist for the EU FP5 project MUMIS (Multimedia Indexing and Searching). He has also worked at the University of Saarland, conducting 2 projects, one on a linguistic infrastructure for eContent and the other one, Esperonto, on the relation between NLP and Semantic Web. He was till the end of 2008 leading the DFKI contribution to the European Network of Excellence "K-Space" (Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content) and he was coordinating the research work-packages of the IP MUSING (MUlti-industry, Semantic-based next generation business INtelliGence). He was also involved in the German national project Theseus, more specifically within the MEDICO sub-project (http://www.theseus-programm.de/en/920.php). In the 7th Framework Programme, Thierry Declerck is involved in the Monnet project (Multilingual Ontologies for Networked Knowledge, http://www.monnet-project.eu). And he is the coordinator of the TrendMiner project, which started in November 2011 (www.trendminer-project.eu). Thierry Declerck is also actively involved in standardisation activities at both ISO TC37/SC4 and W3C.

Massimo Romanelli is a senior consultant at the German Research Centre for Artificial Intelligence (DFKI GmbH). He holds a Computational Linguistics diploma from the University of Saarland, Saarbruecken and joined DFKI in April 2000. At DFKI he is responsible for international standards and ontologies at the Intelligence User Interfaces Department (IUI). He worked on several publicly and industry funded projects (e.g., SMARTKOM, SMARTWEB, THESEUS, VALUE-IT) and published various articles in the fields of semantic web and were involved in the organisation of various events (e.g., SAMT 2010, FOIS 2008, W3C MMI/VB f2f Meeting 2008).

Currently, he is Advisory Committee Representive for DFKI at the World Wide Web Consortium (W3C) where he is also active as member of the Multimodal Interaction Working Group (W3C MMI WG) and Unified Service Description Language Incubator Group (W3C USDL XG). His scientific activity focus on dialog systems knowledge based interaction with content (see the IKS (http://www.iks-project.eu/)).