

Guest Editorial: Content Analysis and Indexing for Advanced Multimedia Services

Alberto Messina · Andrea Basso · Werner Bailer

Published online: 19 March 2015

© Springer Science+Business Media New York 2015

The transition to digital broadcasting and the concomitant rise of new media channels has meant a significant increase in communication potential for media companies, which can now leverage the advantages of online digital technologies to increase the value and attractiveness of their services, thus gaining renewed value from content.

A side effect of such abundance of content is that consumers are overwhelmed with “information overload”. In fact, while digital and Internet services are in principle more appealing due to the opportunity they offer to increase the number of thematic channels, the richness of distributed content and the possibility for the users to interact, accessibility of and interaction with such content still remain mostly unresolved problems. On the media production side, professionals often experience dual problems in content selection and organization for cross-media and interactive productions. The organization of content into searchable units through the use of flexible and scalable indexing techniques is seen as one solution to these problems. In addition, it is of paramount importance to develop the ability to generate, represent and distribute such informational units (e.g., indexes) in a way that is consumable and manageable by a wide range of end user terminals, and seamlessly integrated with web services and mobile apps.

In response to these challenges this special issue presents a number of research works that collectively address many of the aforementioned aspects and enlighten several additional opportunities.

The first two contributions are in the direction of advancing the state of the art in news content indexing with the specific objective to detect and retrieve relevant conceptual entities. Authors of “Multi-modal fusion for associated news story retrieval” [6] investigate multi-modal approaches to retrieve associated news stories sharing the same main topic. In the visual domain this is done through near duplicate keyframe/scene detection based on local signatures

A. Messina (✉)

RAI-Radiotelevisione Italiana, Centre for Research and Technological Innovation, Via Cavalli 6,
10138 Turin, Italy
e-mail: a.messina@rai.it

A. Basso

MI. TO. Technology, Milan, Italy

W. Bailer

Joanneum Research, Digital, Institute for Information and Communication Technologies, Graz, Austria
e-mail: werner.bailer@joanneum.at

to identify stories with mutual visual cues. In addition, they develop a semantic signature that contains pre-defined semantic visual concepts in a news story and visually combine local and semantic signature similarities to obtain enhanced visual content similarity. In the textual domain, Automatic Speech Recognition (ASR) and refined Optical Character Recognition (OCR) are used to enhance textual similarity. They finally use a mixed early/late fusion paradigm to boost the retrieval performance. Experimental results show the usefulness of the enhanced visual content similarity and the early fusion approach, and the superiority of the late fusion approach. Still in the news domain, the work presented in “Automatic discovery of person-related named-entity in news articles based on verb analysis”, [1] starts from the simple empirical observation that in news articles there is always at least one verb attached to the person(s) mentioned in the news to demonstrate the hypothesis that there must exist some verbs that specifically describe human being conducts within a news article. Authors develop an approach which aims to identify named-entity (NE) that performs human activity automatically by studying the nature of the verb associated with human activity via TreeTagger, Stanford packages and WordNet. The experimental results show that it is viable to use a verb in identifying “person name” entity type. The approach is applicable also to small text size articles and does not require training data set and anaphora resolution.

In “Joint utilization of local appearance and geometric invariants for 3D object recognition” [3] authors introduce a novel method for 3D object recognition, which utilizes well-known local features in a more efficient way, without any reliance on partial or global planarity. Geometrically consistent local features, which form the crucial basis for object recognition, are identified using affine 3D geometric invariants. The utilization of 3D geometric invariants replaces the classical 2D affine transform estimation/verification step, and provides the ability to directly verify 3D geometric consistency. The main contribution of the proposed approach lies in this ability of incorporating highly discriminative affine invariant 3D information much earlier in the process of matching in comparison with its counterparts. The accuracy and robustness of the method in highly cluttered scenes, without any prior segmentation or post 3D reconstruction requirements, are presented in the experiments.

Current approaches at audiovisual digital libraries annotation usually adopt metadata harvesting to build a centralized index from separate digital libraries. This approach usually suffers from the problem of metadata inconsistency. To overcome this issue authors of “CoBITs: A distributed indexing approach to collaborative content-based multimedia retrieval across digital archives”, [4] start from the assumption that distributed crawler-based approaches can simplify the design of indexing and query processing steps by maintaining the data to be indexed local to the machine for crawling. To reduce the loads in each archive, they dynamically distribute the tasks of crawling, indexing, and query processing depending on the response time. The main objective of the study is to prove the potential of the proposed approach in load balancing with appropriate task distribution.

Consumer devices are increasingly becoming useful tools also in professional media production, and two of the research works of this special issue specifically address this topic, respectively on the production and the consumption side.

The work presented in, “Real-time selection of video streams for live TV broadcasting based on Query-by-Example using a 3D model” [7] addresses one of the most critical aspects of multisource video production and how this can be cost-effective using consumer shooting devices. In such environments Technical Directors have the issue of being able to search for interesting shots (e.g., a certain view of a specific car in a race) among many video sources in real-time, many of which can be consumer-level cameras. The authors describe a system helping technical directors visually define interesting sample views of one or more objects by re-creating cameras’ views in a 3D engine, and apply 3D geometric computations to enable an

efficient and precise real-time selection. The system is based on the usage of a similarity measure to rank the candidate cameras.

The availability of robust and cheap gesture tracking equipment on the user's side enables the development of novel mechanisms to engage users with broadcasted content. The work described in "Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware" [5] is about the design and development of a robust marker-less hand/finger tracking and gesture recognition system using low-cost hardware. Robust and fast hand tracking despite complex background and motion blur is achieved by translating the detected hands or gestures into different functional inputs and interfaces with other applications via several methods. The results show that an intuitive HCI and motion gaming system can be achieved with minimum hardware requirements.

Expanding and enhancing the user experience is among the challenges of modern media companies. This is the subject of "Object-based audio for interactive football broadcast", [2], in which authors describe the development of real time audio object event detection and localisation, scene modelling and processing methods for multimedia data, which allow users to navigate the event by creating their own unique user-defined scene. As part of the first implementation of the system a test shoot was carried out capturing a live Premier League football game and methods have been developed to detect, analyse, extract and localise salient audio events from a range of sensors and represent them within an audio scene in order to allow free navigation within the scene.

Overall, this special issue presents a good overview of some of the current challenges in content analysis and indexing for advanced multimedia services, providing a set of research articles that addresses many issues in the end-to-end value chain from producers to consumers.

References

1. Goh H-N, Soon L-K, Haw S-C, "Automatic discovery of person-related named-entity in news articles based on verb analysis", DOI [10.1007/s11042-013-1618-2](https://doi.org/10.1007/s11042-013-1618-2)
2. Oldfield R, Shirley BG, "Object-based audio for interactive football broadcast", DOI [10.1007/s11042-013-1472-2](https://doi.org/10.1007/s11042-013-1472-2)
3. Soysal M, Alatan AA, "Joint utilization of local appearance and geometric invariants for 3D object recognition", DOI [10.1007/s11042-013-1622-6](https://doi.org/10.1007/s11042-013-1622-6)
4. Wang J-H, Chang H-C, "CoBITs: A distributed indexing approach to collaborative content-based multimedia retrieval across digital archives", DOI [10.1007/s11042-013-1461-5](https://doi.org/10.1007/s11042-013-1461-5)
5. Yeo H-S, Lee B-G, Lim H, "Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware", DOI [10.1007/s11042-013-1501-1](https://doi.org/10.1007/s11042-013-1501-1)
6. Younessian E, Rajan D, "Multi-modal fusion for associated news story retrieval", DOI [10.1007/s11042-013-1404-1](https://doi.org/10.1007/s11042-013-1404-1)
7. Yus R, Ilarri S, Mena E, "Real-time selection of video streams for live TV broadcasting based on Query-by-Example using a 3D model", DOI [10.1007/s11042-013-1550-5](https://doi.org/10.1007/s11042-013-1550-5)

Alberto Messina is R&D Area Coordinator at RAI, in the area concerning technologies for automatic extraction of information from media. He is co-author of more than 80 publications and collaborates with international research institutions worldwide. He is chair of the EBU Strategic Programme on Media Information Management. He worked in many financed projects including VISION Cloud, winner of IBC 2013 Special Award. His research activities have been awarded in 2010 with the award of Confindustria ICMT, delivered to RAI by Giorgio Napolitano, President of Italian Republic.