Editorial
Knowledge-based digital media processing

It is now broadly acknowledged within the research community and beyond, that the development of enabling technologies for new forms of self-adaptive, self-organising, context and user aware content-based applications requires a targeted confluence of knowledge, semantics and low-level media processing. Such technology is underpinned by systems that are increasingly complex and it is key in many applications including interactive media, forensics, biometrics, medical imaging, and multimedia search and retrieval. The latter is a crucial application since the exponential growth of audiovisual data, along with the critical lack of tools to record that data in a well-structured form, is rendering useless vast portions of available content. Clearly, content is worthless if it cannot be found and used. As a consequence multimedia information retrieval technology is destined to become pervasive in almost every aspect of daily life and a pillar for key achievements in future scientific and technological developments.

Although the fundaments of information retrieval were laid many years ago, this was done for text databases and most of current well-established retrieval tools are only suitable for text mining. Compared to text-based information retrieval, image and video retrieval is not only less advanced but also more challenging. Though writing was explicitly developed to share and preserve information while pictures and sound have been traditionally used to express human artistic and creative capacity, this model is changing in the digital age. This shift is also revolutionising the way people process and look for information: from text only to multimedia-based search and retrieval. This progression is a consequence of the rapid growth in consumer-oriented electronic technologies; for example, digital cameras, camcorders and mobile phones, along with the expansion and globalisation of networking facilities. Indeed, the availability of a wide range of digital recorders accessible to anyone, from cheap digital cameras to complex professional movie capturing devices, is enabling the wide use of images, diagrams and other audiovisual means to record information and knowledge while boosting the content growth in digital libraries. The immediate consequence of this trend is that generating digital content has become easy and cheap while managing and structuring it to produce effective services has not. This applies to the whole range of content owners, from professional digital libraries with their terabytes of visual content to the private collector of digital pictures stored in the disks of conventional personal computers.

To get closer to the vision of useful multimedia-based search and retrieval, the annotation and search technologies to be employed need to be efficient and use semantic concepts that are natural to the user. To develop the technology able to produce accurate levels of abstraction in order to annotate and retrieve content using queries that are natural to humans is the achievement needed to bridge the gap between low-level features or content descriptors that can be computed automatically, and the richness and subjectivity of semantics in user queries and high-level human interpretations of audiovisual media. To bridge this gap is a challenge that has captured great attention from researchers in computer vision, pattern recognition, image processing and other related fields, evidencing the difficulty and importance of such technology and the fact that the problem is unsolved. This challenge offers the possibility of adding the audiovisual dimension to well-established text databases to build multimedia-enabled information retrieval. If the goal is to retrieve audiovisual content using semantic structures – for example, words or sentences, which are natural to humans – two profound challenges become evident: how to deal with the subjective interpretation of images by different users under different conditions; and how to link the semantic-based query with low-level metadata. The first problem originates in the fact that perceptual similarity is user and context dependent. The second challenge is a synonym of the semantic gap.

This Special Section presents work on integrative research aimed at low-level analysis, classification and semantic-based structuring of digital media. Most of the work presented in this section has originated in two large international cooperative projects funded by the European Commission under the sixth framework programme of the Information Society Technology: aceMedia and COST292. The mandate and scope of these two projects is very generic, embracing several applications that rely on technology for bridging the gap.

A total of eight papers were selected for this Special Section. The first four papers combine low and high-level analysis to extract semantics from visual and multimedia data. In their paper Petridis *et al*. present a knowledge infrastructure and an experimentation platform for semantic annotation to narrow the semantic gap. The used multimedia ontology is extended and enriched to include low-level audiovisual features and descriptors. The authors construct an ontology that includes typical instances of high-level domain concepts together with a formal specification of the corresponding low-level visual descriptors. This model is exploited by a knowledge-assisted analysis framework that handle problems like segmentation, tracking, feature extraction and matching in order to automatically create the associated semantic metadata. The second paper by Dorado, Djordjevic, Pedrycz and Izquierdo addresses the problem of efficient image selection for concept learning. The proposed model exploits the capability of support vector classifiers to learn from relatively small number of patterns. The approach uses unsupervised learning to organise images based on low-level similarity in order to assist a professional annotator in picking positive and negative samples for a given concept. Active learning is then used to refine the classification results. The paper by Meessen, Xu and Macq targets semiautomatic semantic extraction from long video sequences. They introduce a software platform for remote and interactive browsing of summaries of long video sequences and extraction of semantic links between shots and scenes in their temporal context. In the fourth paper Kiranyaz and Gabbouj argue that from the point of view of content-based multimedia

retrieval, audio cues can be even more important than their visual counterparts. They present a generic audio based multimedia indexing and retrieval framework. This framework supports the dynamic integration of the audio feature extraction modules during the indexing and retrieval phases and therefore, provides a test-bed platform for developing robust and efficient aural feature extraction techniques. The design of the framework is based on the high-level content classification and segmentation in order to improve the speed and accuracy of the aural retrievals.

Clustering and automatic image classification is crucial for keyword based annotation. Two papers address this specific problem. The paper by Grira, Crucianu and Boujemaa presents an image classification approach that exploits pairwise constraints between images. A semi-supervised clustering algorithm using the underlying pairwise-constrained competitive agglomeration and based on a fuzzy cost function is introduced. Burghardt and Ćalić describe in the next paper an algorithm to categorise animal locomotive behaviour. It combines detection and tracking of animal faces in wildlife videos. The detection algorithm is based on a human face detection method, utilising Haar-like features and AdaBoost classifiers. The face tracking module uses a specific interest model that combines low-level feature tracking and the detection algorithm. The annotation classes of locomotive processes for a given animal species are predefined by a large semantic taxonomy on the wildlife domain.

The last part of the Special Section contains two papers targeting automatic object segmentation in images, as a fundamental step towards object recognition for image annotation. The paper by Pratikakis *et al.* describes a method for unsupervised content-based image retrieval. It is based on a meaningful segmentation procedure that can provide proper distributions for matching via the Earth mover's distance as a similarity metric. The segmentation procedure is based on a hierarchical watershed-driven algorithm that extracts semantically meaningful image regions automatically. Key features of the techniques are the proposed many-to-many region matching and the region weighting strategy to enhance feature association and discrimination. In the second paper Lu *et al.* recognise through subjective tests that tree based representations are amenable to segmentation based on simple descriptions. The method considers the novel application of relative topology of constituent regions within the tree representation and applies it to form registrations in the spatial and temporal domains. To achieve this requires the use of graph matching techniques to circumvent the non-direct mapping between tree conjugates and derived models. Search strategies are successfully applied in the tree using multiple hypotheses testing in a Bayesian formulation framework facilitating object registration from very simple models.

This Special Section has assembled a small sample papers originating from well-known research institutions across Europe. The contributing authors were instrumental in the completion of the Special Section and I would like to thank all of them. The anonymous referees also played a crucial role in the review and selection process ensuring the Special Section includes only the submissions of the highest technical quality. Finally, I would also like to thank the IEE editorial team, especially Keith Martin, for their help and very efficient handling of this Special Section.

EBROUL IZQUIERDO

IEE Proceedings online no. 20069007
doi:10.1049/ip-vis:20069007

Ebroul Izquierdo is Chair of Multimedia and Computer Vision and head of the Multimedia and Vision Lab at Queen Mary, University of London. Prof. Izquierdo received the Dr. Rerun Naturalium (PhD) from the Humboldt University, Berlin, Germany, in 1993. From 1993 to 1997 he was with the Heinrich-Hertz Institute for Communication Technology (HHI), Berlin, Germany, as associated researcher. From 1998 to 1999 Dr. Izquierdo was with the Department of Electronic Systems Engineering of the University of Essex as a senior research officer. Since 2000 he has been with the Electronic Engineering Department, Queen Mary, University of London. Prof. Izquierdo was the UK representative of the EU Action Cost211 and coordinated the EU IST project BUSMAN. Currently, he is a main contributor to the IST integrated projects aceMedia and MESH. He also coordinates the EU Action Cost292 and the FP6 network of excellence on semantic inference for automatic annotation and retrieval of multimedia content, K-Space. Prof. Izquierdo is associate editor of *IEEE Transactions on Circuits and Systems for Video Technology* and has served as guest editor of three Special Issues of the IEEE TCSVT and one Special Issue of the journal *Signal Processing: Image Communication*. He is a Chartered Engineer, a senior member of the IEEE, the IEE and the British Machine Vision Association. He is member of the programme committee of the IEEE Conference on Information Visualization, the international program committee of EURASIP & IEEE Conference on Video Processing and Multimedia Communication and the European Workshop on Image Analysis for Multimedia Interactive Services. Prof. Izquierdo has served as session chair and organiser of invited sessions at several conferences. He has been the chair of the European Workshop on Image Analysis for Multimedia Interactive Services, London 2003 and Seoul 2006, the European Workshop for the integration of Knowledge, Semantics and Content, London 2004 and 2005 and the Mobile Multimedia Communications Conference MobiMedia 2006. He has published over 150 technical papers and book chapters.