Knowledge-based Semantic Annotation and Retrieval of Multimedia Content

Giorgos Akrivas, Georgios Th. Papadopoulos, Matthijs Douze, Johannes Heinecke, Noel O'Connor, Carsten Saathoff, Simon Waddington

***Abstract*— aceMedia is a 4 year EC part-funded FP6 Integrated Project, ending in December 2007. The project has developed tools to enable users to manage and share both personal and purchased content across PC, STB and mobile platforms. Knowledge-based analysis and ontologies have been successfully exploited in an end-to-end system to enable automated semantic annotation and retrieval of multimedia content. The paper briefly describes the objectives of aceMedia and the application of knowledge-based analysis in the project.**

***Index Terms*—multimedia content analysis, ontologies**

I. Introduction

The explosion in the range and capability of devices, such as camera phones, that can capture digital images and videos, has resulted in users being able to rapidly create large collections of multimedia content. At the same time, the increased connectivity of devices, via personal, local and wide area networks, has resulted in a rapid growth of technologies that allow users to purchase, share and exchange content. However, the lack of effective tools for annotating and organising content has, to date, made it difficult for users to manage their content collections. This problem has been addressed by the aceMedia EC FP6 Integrated Project.

Manuscript received 15th October, 2007. This research was supported by the EC IST FP6 project aceMedia (http://www.acemedia.org) under contract FP6-001765.

G. Akrivas is with NTUA, Athens, Greece (e-mail: Gakrivas@image.ntua.gr).

G. Th. Papadopoulos is with ITI, Thessaloniki, Greece (e-mail: papad@iti.gr).

M. Douze is with INRIA, Grenoble, France (email: Matthijs.Douze@inrialpes.fr).

J. Heinecke is with France Telecom R&D, Lannion, France (email: johannes.heinecke@orange-ftgroup.com).

N. O'Connor is with Dublin City University, Ireland (email: oconnorn@eeng.dcu.ie).

C. Saathoff is with University of Koblenz, Germany (email: saathoff@uni-koblenz.de).

S. Waddington is with Motorola Ltd, Basingstoke, UK (email: simon.waddington@motorola.com).

II. Semantic Content Analysis in aceMedia

aceMedia has developed a set of visual and textual content analysis modules to enable automatic generation of semantic metadata, examples of which are described in the following section.

The semantic visual analysis techniques for images and video make use of implementations of low-level feature analysis in the aceToolBox [3]. These are based on a subset of the visual descriptors defined in the MPEG-7 standard. The work carried out in aceMedia can be seen as a contribution to bridging the semantic gap between low-level information that can easily be generated by computers and high-level semantic information that is understandable by humans.

A fundamental concept developed in aceMedia is that of the *Autonomous Content Entity* (ACE) [6], a multimedia object containing the content itself, the metadata and an intelligence layer. The aim of the intelligence layer is to assist the user in the complex task of multimedia content management by enabling the content items to perform actions on behalf of the user, independent of where they reside. The intelligence layer enables such features as self-organisation, self-governance, self-adaptation and personalised retrieval of content [5]. These intelligent features are to a large extent reliant on the quality of the semantic metadata generated.

The aceMedia framework and content analysis algorithms have been implemented in C++ on a PC platform. A mobile prototype based on a Motorola A910 handset has been developed which allow users to upload captured content to the aceMedia PC server for processing, as well as to perform intelligent search and retrieval, and organise their content collections.

III. Applications of Knowledge-based Analysis

In aceMedia, two separate applications were developed based on the aceMedia analysis modules. The first was aimed at personal usage (PCS) and considered the Holiday and Family domains. The second was aimed at professional users (CCM) and considered motor sports and tennis content.

A. Knowledge Assisted Analysis of Images and Video

aceMedia developed solutions for the automatic semantic annotation of both images and video. Support Vector Machines proved to be a robust methodology for discovering complex relationships between the numerical image data and the perceptually higher-level concepts. Use is made of domain-specific spatial knowledge, which exploits the spatial context

of the objects within an image, in order to discriminate between objects with similar visual characteristics [7]. Semantic analysis of video content is very challenging, due to the large amount of information present and the complexity of the relationship between low-level data (e.g. pixel luminance, colour, motion vectors) and higher-level concepts. An example of event detection that was performed, based on Hidden Markov Models, was to detect the occurrence of events *Serve*, *Rally*, *Break* and *Replay* in tennis match videos.

B. Person Detection and Identification

Detecting humans in video streams is a challenging problem owing to variations in pose, body shape, appearance, clothing, illumination and background clutter. The objective was to develop a detector that could be used to analyse film and TV content in which the camera and the background are not static. The main challenge was to find a set of features that characterise human motion well, while remaining resistant to camera and background motion.

C. Face Detection and Recognition

Face processing is a very hot research topic due to the number of potential applications. Many difficulties arise in considering unconstrained scenes such as illumination, facial pose and expressions. In particular, previous studies have shown that the performance of face recognition techniques drastically degrades when faces are not in frontal pose. Experimental evaluation of the aceMedia approach has shown it to be very robust to occlusion and noise, as well as to image quality [1], under difficult lighting conditions and with partial occlusions.

D. Visual Concept Detection (VCD)

This work dealt with the classification of natural scenes that are likely to be captured in everyday use of a digital camera. The low-level features that must be extracted from natural images were determined, in order to automatically group them into semantically meaningful categories. VCD tasks such as indoor/outdoor or man-made/natural were considered. Two alternative VCD methods were developed in aceMedia, one of which used global image attributes and a global classifier, and the second which used local patches that are classified individually.

E. Ontological Text Analysis

There are two main applications of ontological text analysis in aceMedia. Manual textual annotations (words, short phrases or whole sentences) are semantically analysed and transformed into an ontological representation which is based upon the ontological vocabulary as defined in the domain ontologies. The current format is RDFS-based. The second task was to translate (textual) user queries into an ontological query language (SPARQL). Further details about the construction of the mappings between domain ontologies and semantic data can be found in [2].

F. Multimedia Reasoning Framework

A multimedia reasoning framework is used to integrate the descriptions produced by the content analysis modules into a final, semantically coherent annotation [4]. This means that possible contradictions in the annotations can be resolved across the different modalities (visual and textual) and the different annotation levels (regions and scenes within an image). A further purpose of the reasoner is to generate enhanced annotations based on those extracted by the content analysis modules. This is done by exploiting the logical relations and associations that characterise the chosen domains, which are encoded in the ontology infrastructure.

IV. Conclusion

The aceMedia project has successfully created a set of integrated tools for performing automatic semantic annotation of image and video content based on the use of ontologies. Finding a general solution to bridging the semantic gap is a long-term research challenge. However, the aceMedia project has demonstrated the viability of practical systems that provide automated human understandable annotations for multimedia content. Automatic semantic annotation is a key enabler for search and retrieval, sharing and privacy management of content for both personal and professional users.

Acknowledgement

The authors acknowledge the contributions of their many aceMedia colleagues in carrying out this research.

References

[1] C. Garcia, M. Delakis. Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(11):1408--1423, 2004.

[2] Marie Chagnoux and Johannes Heinecke: *Aligner ontologies et langues naturelles : gérer la synonymie.* Paper accepted at AFIA 2007, Grenoble 2007.

[3] O'Connor, Cooke, Le Borgne, Blighe, Adamek, The aceToolbox: low-level audiovisual feature extraction for retrieval and classification, 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology (EWIMT'05), London, U.K., 30 November- 1 December 2005.

[4] S. Dasiopoulou, J. Heinecke, C. Saathoff and M. Strintzis, Multimedia reasoning with natural language support. In: Proceedings of IEEE-International Conference on Semantic Computing 2007, Irvine CA, pp 413-420.

[5] D. Vallet, P. Castells, M. Fernández, Ph. Mylonas, and Y. Avrithis. *Personalized Content Retrieval in Context Using Ontological Knowledge.* IEEE Transactions on Circuits and Systems for Video Technology 17(3), March 2007, special issue on the convergence of knowledge engineering, semantics and signal processing in audiovisual information retrieval, pp. 336-346.

[6] A. Matellanes, A. Evans, B. Erdal, "Creating an application for automatic annotation of images and video" in *SWAMM 2006, Edinburgh, UK*.

[7] G. Th. Papadopoulos, V. Mezaris, S. Dasiopoulou and I. Kompatsiaris: "Semantic Image Analysis Using a Learning Approach and Spatial Context", 1st International Conference on Semantics and Digital Media Technologies (SAMT, formely EWIMT), special session on Context-based Multimedia Analysis and Representation, Athens, Greece, December 2006.