ACHIEVING INTEGRATION OF KNOWLEDGE AND CONTENT TECHNOLOGIES: THE ACEMEDIA PROJECT

YIANNIS KOMPATSIARIS AND YANNIS AVRITHIS

Informatics and Telematics Institute
1st Km Thermi-Panorama Rd
570 01 Thermi-Thessaloniki, Greece
{ikom, iavr}@iti.gr

PAOLA HOBSON, TONY MAY AND JOLANDA TROMP

Motorola Ltd, Jays Close, Viables Industrial Estate
Basingstoke, RG22 4PD UK
paola.hobson@motorola.com

Integration of knowledge and multimedia content technologies is important for the future of European industry and commerce. aceMedia is an IST FP6 project which aims to unite these two established disciplines to achieve significant advances by the combination of the two domains. This paper describes research in content processing and knowledge assisted multimedia analysis within the aceMedia project, and provides examples of use which illustrate the benefits of this combined approach.

1 Introduction

While the objective of integrating research into knowledge and multimedia content technologies is important for the future of European industry and commerce, in practice, it is difficult to truly unite two established disciplines without an overarching motivation [1]. The aceMedia project provides exactly the framework needed to enable advances in research into knowledge and semantics and multimedia content processing independently, as well as significant advances by the combination of the two domains.

User frustration when dealing with multimedia content is one reason why take-up of digital content services is not as rapid as may be expected from the business predictions of experts a few years ago. Users are often unable to find the content they want e.g. where web searches return thousands of hits, none of which fits the user's mental model of the target, or they have reduced enjoyment of their digital content collections due to the difficulties in indexing and cataloguing all their acquisitions [2].

The aceMedia project [3] aims to bring together the collective expertise of multimedia and knowledge technology experts to drive research on next generation tools and management for multimedia content, which will establish the EU as a leader in this field. In order to simplify the user experience, the aceMedia project focuses its efforts on knowledge discovery and self-adaptability embedded into media content, which will allow it to be self organising, self annotating, and more readily searched and communicated.

Users in the future will access multimedia content using a variety of devices such as mobile phones and set-top-boxes, as well as via broadband cable or wireless to their PC. Of key importance to user satisfaction is delivery and presentation of the chosen content in a form which matches the capability of the user device. aceMedia will use knowledge about user preferences and current context (e.g. their device, available bandwidth, etc) in order to adapt the multimedia content being delivered to the user [4]. This is made most efficient by the use of emerging scalable video codecs.

In this paper, first the aceMedia system overview is presented, followed by an outline of the scalable video coding research and a summary of the key knowledge engineering advancements which will be used to drive media adaptation. The integration of these two research strands is illustrated, in the context of a user-centred example, illustrating the benefits of this combined approach.

2 aceMedia system overview

The aceMedia integrated project draws together fundamental research in knowledge technologies and multimedia processing, within a user centred design framework.

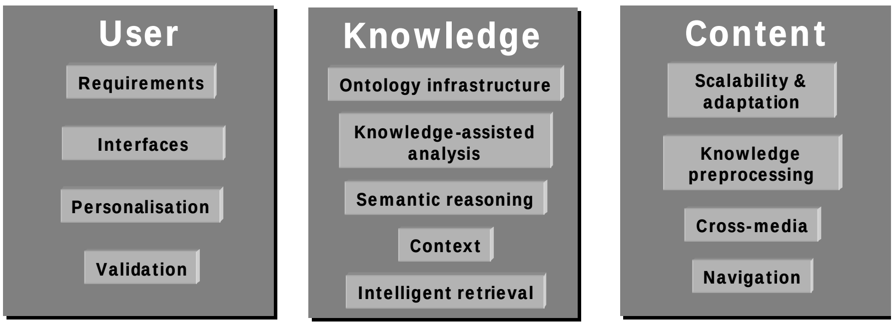

Figure 1. aceMedia project overview.

As is depicted in Figure 1, the main workplan of aceMedia is built around the outcomes of the core research workpackages of aceMedia: WP2 (User), WP3 (Content) and WP4 (Knowledge). The content subsystem interacts with multimedia content obtained from the content creator and handles the essential processing tasks required by the content provider, service provider and network operator, including pre-processing, scalable coding, cross-media adaptation and visualization. The user subsystem interacts with the aceMedia end user and supports the elementary operations required by the service provider and device manufacturer, including content search, browsing personalisation and consumption. Finally, the knowledge subsystem interacts with aceMedia knowledge to implement all intermediate steps of intelligent analysis and reasoning throughout the aceMedia value chain. This builds the foundation of all knowledge-driven operations of aceMedia, including the components of knowledge-assisted multimedia analysis, semantic reasoning and content search/retrieval.

Some of these steps take place during production of the fundamental element of aceMedia , the Autonomous Content Entity (ACE production) e.g. pre-processing, scalable coding, and knowledge-assisted analysis, others during ACE usage, e.g. content search/retrieval, navigation/rendering and personalisation, while others during both e.g. semantic reasoning.

These technical workpackages are supported by the WP2 (User requirements), WP5 (System specification), WP6 (Integration and applications) workpackages and by appropriate management, exploitation and dissemination activities.

3 Scalable coding for media adaptation

3.1 Content adaptation

The heart of the aceMedia system is the ACE - an Autonomous Content Entity which can adapt itself to meet the requirements of the desired application and user profile. The ACE consists of content, metadata, and an intelligence layer which can act on the content and/or metadata to perform actions autonomously, which enable the user to achieve their desired goals. Content adaptability according to available bandwidth, is at the heart of the ACE concept, at the media creation stage. The objective is to provide tools which allow the content to be encoded just once, but used in a wide range of contexts, where the encoded bitstream may be adapted to suit available bandwidth and/or device capability. In addition to fully scalable codecs, scalable metadata, and middleware solutions supporting enhanced portability across networks and platforms are required.

Figure 2 shows the adaptation architecture for the aceMedia system. The content in the ACE will be adapted in response to user preferences, environmental conditions, network resources and terminal capabilities before being delivered to the terminal. The media will be described using standards such as MPEG-7 MDS (Multimedia Description Schemes) [5] and RDF (Resource Description Framework) [6], allowing the knowledge about the content to be easily accessed. The syntax of the encoded content will be described using the MPEG-21 DIA standard [7], allowing transformation of the content without having to fully decode the content. Media Adaptation Hints which are part of the MPEG-7 content descriptors standard will be used to ensure that the content is adapted optimally.

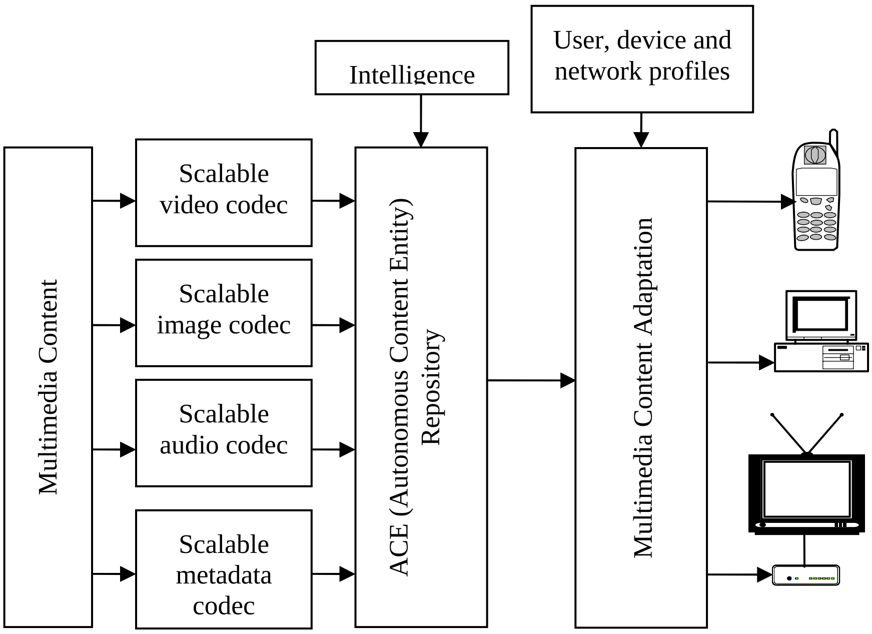

Figure 2. aceMedia adaptation architecture.

The adaptation will reduce the bit rate of the content to target different network types and conditions, and may also require the content to undergo a modality adaptation, e.g. text to speech. The adaptation behaviour will take account of multiple related streams (e.g. 64 kbps audio with 300 kbps video). This functionality is required in the process of adaptation to user profiles, e.g. hearing impaired people may prefer high visual quality and low audio quality streams (or no audio at all, or closed captions, etc).

Although the adaptation functionality is shown as a single box in the diagram, the adaptation may be distributed throughout the network. The initial adaptation carried out with the best information available at the content repository can be complemented by having additional adaptation nodes at key points in the network which are better placed to be able to optimally adapt the stream. If the same data is being multi-casted to several terminals then a distributed adaptation architecture will be able to make best use of network resources; a repository based adaptation engine on its own would have to generate multiple simulcast content streams.

Content adaptation is key to realising the value of multimedia assets. For example, the owner of footage from last year's national football championships may wish to make archive clips available on a pay-per-view basis to mobile phone and home PC users, as well as on a multiple usage basis to professional content aggregators. The aim of aceMedia is to provide media codecs which allow this content to be encoded only once, e.g. using scalable codecs, with adaptation to suit each user and their prevailing device and network conditions. Therefore, the professional user would have access to the full resolution footage, the home PC user using broadband would receive and decode a reduced bandwidth stream, and the aceMedia system would select appropriate portions of the bitstream to send to the mobile phone user, according to the users' personal preferences, as captured in each user profile. Allowing each type of user to receive the content in the format of their choice greatly increases the value of that content, and therefore augments the premium to be paid by customers.

3.2 Scalable video codec development

The video adaptation functionality required for aceMedia can be achieved with a video codec that satisfies the following requirements:

- Produces a single bit stream that can be adapted to several levels of spatial or temporal resolution.
- Produces a single bit stream whose bit rate can be reduced by reducing its spatial or temporal resolution, or its visual quality, without decoding and re-encoding the stream.
- Has a computational complexity that decreases proportionally to the video resolution encoded or decoded.
- Produces a single bit stream that can be randomly accessed at any spatial, temporal and quality resolution, and that permits multiple successive extractions of lower quality and resolution bit streams.
- Produces a single bit stream .

The scalable technologies so far adopted in current standards are limited. In particular, present solutions to scalable coding, which involve closed loop coders, cannot create fully embedded bit streams without seriously degrading the efficiency of the video coding.

Fully embedded bit streams require entirely new technology such as open loop wavelet based codecs. Some implementations of open loop wavelet codecs have been shown to have comparable performance to present-day single layer coders (MPEG-4 AVC) while offering scalable functionality. However, these codecs tend not to perform well at low bit rates, and current implementations have yet to achieve real time implementations.

In order to meet the wide ranging scalability requirements of aceMedia the project is investigating the fusion of the closed loop techniques used in MPEG-4 AVC that provide high coding efficiency at low bit rates, frame rates, and spatial resolutions, and the open-loop, wavelet based techniques that provide high coding efficiency at high bit rates, frame rates and spatial resolutions.

4 Knowledge assisted multimedia analysis

In addition to content adaptation, another R&D objective of aceMedia is to develop the necessary infrastructure, technology and tools for knowledge representation, extraction and usage throughout the aceMedia value chain. This technology will permit structured knowledge representation in aceMedia applications, provide tools to automatically analyse ACE content, extract knowledge and generate ACE metadata and annotation, as well as support intelligent aceMedia content search and retrieval services.

Intense past research in the domains of knowledge representation and reasoning with knowledge has, over the last decade, gained new interest in the context of the Semantic Web. New languages such as RDFS (Resource Description Framework Schema) [8] and OWL (Web Ontology Language) [9] have been defined by the World Wide Web consortium (W3C) in order to render meaning to information on the web and allow for better methods of search and retrieval. As a next step, inference rules and logic are to be used by intelligent applications to derive new information from existing information on the web. Ontologies, which define a set of meanings for a specific domain of information, play an important role in the implementation of the Semantic Web [8]. Following this approach, aceMedia has developed novel ontology structures in terms of both methodology and expressiveness in order to address the additional requirements of multimedia resources. More specifically, the developed infrastructure consists of the core ontology based on extensions of the DOLCE core ontology [11] and the multimedia-specific infrastructure components. These are, the Visual Descriptor Ontology (VDO), which is based on an RDFS representation of the MPEG-7 Visual Descriptors [12]and the Multimedia Structure Ontology (MSO), based on the MPEG-7 MDS [5]. Furthermore, the Visual Descriptor Extraction (VDE) tool has been developed to support the construction of domain ontologies enriched with multimedia features.

Within aceMedia, this knowledge infrastructure will be used by applying ontology-based discourse representation, analysis and semantic reasoning on multimedia resources. Therefore, an experimental framework is being developed including methods that automatically segment images, video sequences and key frames into areas corresponding to salient semantic objects (e.g. cars, road, people, field, etc), track these objects over time, and provide a flexible infrastructure for further analysis of their relative motion and interactions, as well as object recognition, metadata generation, indexing and retrieval. This problem can be viewed as relating symbolic terms (concepts of the related domain ontology) to visual information by utilizing syntactic and semantic structure in a way similar to approaches in speech and language processing. More specifically, MPEG-7 compliant low-level multimedia features (e.g. MPEG-7 visual descriptors) are assigned to semantic concepts thus forming an a-priori knowledge base. Processing is then performed by relating high-level symbolic representations to extracted features in the signal (image and temporal feature) domain, thus identifying objects and their relations in the multimedia content. Basing such a representation on an ontology, one can capture both concrete and abstract relationships between salient visual properties.

As an example, consider a user seeking archive material about a relative, who was a minor professional tennis player during the 1970's. The user's family have given her access to hundreds of family photos and hours of home movies collected over the years. aceMedia can been used to annotate the content in a meaningful way, which allows the user to search and group the content according to various criteria e.g. all games played on grass or gravel-type courts.

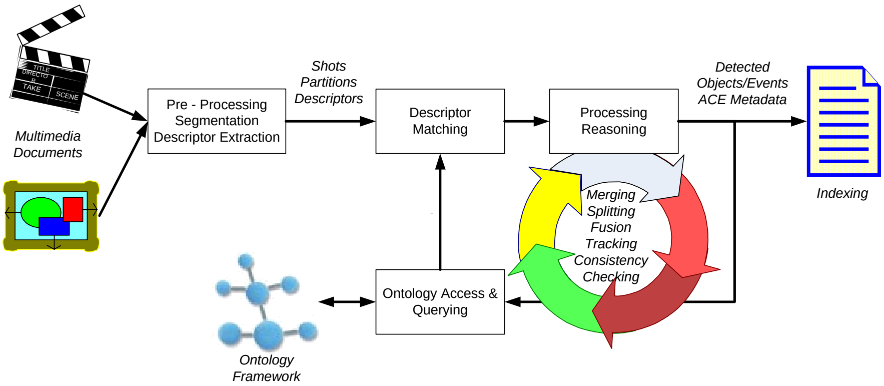

Figure 3. Knowledge-assisted multimedia analysis architecture.

The architecture of knowledge-assisted analysis is shown in Figure 3. A first step of analysis is pre-processing, including spatio-temporal segmentation and visual descriptor extraction. Access to and querying of the ontology framework is then employed for mapping and matching between conceptual and visual description, in order to get an initial identification of image regions. This is then used for the recursive refinement, identification and labelling of the extracted regions and their description. This involves several processing and reasoning tasks including region merging/splitting, image partition fusion, partonomic and spatiotemporal relation processing, tracking consistency checking. The ontology framework involved in this process includes the ontology library (core ontology, VDO, MSO, domain ontologies, mapping rules), ontology middleware (query service, ontology selection, ontology access, administrative services), and reasoning tools (visual descriptor matching, object detection using rules).

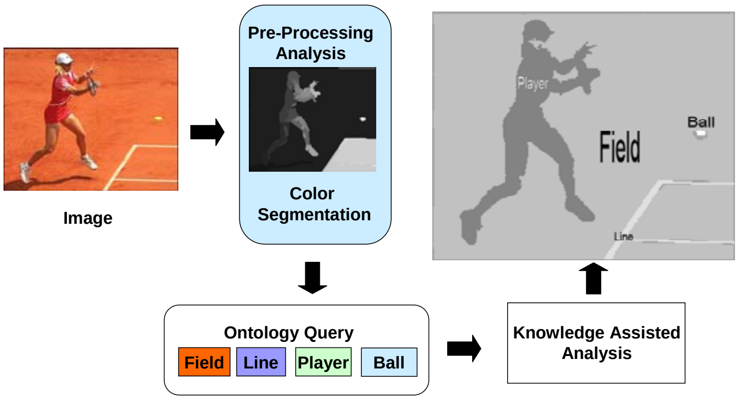

Figure 4. Example of knowledge-assisted multimedia analysis.

An example is shown in Figure 4, where a sequence depicting a tennis player is first pre-processed by means of color segmentation. Visual descriptors of concepts, including “player”, “field”, “line”, “ball”, residing in the multimedia ontology, are selected and employed for the precise fusion of regions, extraction and labelling of all objects in the scene, as shown in the bottom-right of the Figure. Visual descriptors include color, texture and shape.

Several other processes are employed to assist in this challenging analysis task. Ontology learning techniques are employed to automatically train visual concept detectors in order to handle entities that cannot be easily detected through visual descriptor matching, and rapidly identify the context of an entire scene (e.g. indoor/outdoor scene). Specialized tasks are devoted to person detection and identification, including human body/face detection and recognition. Audio features are exploited by means of tools like volume envelope extraction, silence detection and speech/music classification. Finally, ontological text analysis is employed on available textual metadata. All this input is used for the detection of high-level objects and events through discourse analysis and inferencing, assisted by novel context modelling and analysis methodologies. A hierarchical, conceptual representation of content is automatically generated and used to construct the ACE scalable metadata, and finally provide the necessary input for ACE content indexing and retrieval.

Returning to our user assembling a collection of material associated with her tennis playing relative, the aceMedia system can be used to sort and annotate both purchased and home-created content into themes that will reflect the different events and places, types of games (single or double) of their career. In an automatic or semi-automatic (supervised) mode, the aceMedia system would propose different subject groupings based on the content descriptions that aceMedia generates, such as early years’ home photos, minor games, important games, highlights, interviews, advertisements etc. The user may correct incorrect content descriptions or groupings. The aceMedia system learns from the user's actions such that it proposes additional groupings to her based on the visual similarities of her current groupings. Where the content has no associated metadata, the aceMedia system carries out video and audio analysis, driven by a tennis domain ontology which contains lists of information and keywords, enriched with multimedia features to seek semantic meaning in the content.

5 Conclusions

The target of the aceMedia project to integrate knowledge and multimedia content technologies have been outlined in this project, focusing on the benefits of the end user, in the context of a user-centred scenario. The ACE concept will be realised by developing integrated systems and applications that span the complete content value chain from production to consumption by users. Knowledge will play a central role in this process, through tools to automatically analyse content, generate ACE metadata and annotation, and support intelligent content search and retrieval services.

Acknowledgements

aceMedia is supported by the European Commission under contract FP6-001765. aceMedia partners are Fratelli Alinari, Belgavox, Dublin City University, France Télécom R&D, Fraunhofer FIT, INRIA, University of Karlsruhe, CERTH-ITI, Motorola, Philips Electronics Nederland, Queen Mary University of London, Telefónica I+D, Thomson, and Universidad Autonóma de Madrid.

References

[1] S.-F. Chang. The holy grail of content-based media analysis. *IEEE Multimedia*, 9(2):6–10, Apr.-Jun. 2002.

[2] A. Yoshitaka and T. Ichikawa. A survey on content-based retrieval for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, 11(1):81–93, Jan/Feb 1999.

[3] http://www.acemedia.org

[4] M. Davis, N. Good, and R. Sarvas. From Context to Content: Leveraging Context for Mobile Media Metadata. In *Proc. of Int. Conf. on Mobile Systems, Applications, and Services*, Boston, MA, June 2004.

[5] ISO/IEC 15938-5 FCD Information Technology – Multimedia Content Description Interface - Part 5: Multimedia Description Schemes, March 2001, Singapore.

[6] http://www.w3c.org/RDF/

[7] MPEG MDS Subgroup. Multimedia framework – Part 7: Digital item adaptation. Presented at the Awaji MPEG Meeting, Japan, Dec. 2002.

[8] http://www.w3.org/TR/rdf-schema/

[9] http://www.w3c.org/2004/OWL/

[10] Steffen Staab and Rudi Studer. *Handbook on Ontologies*. International Handbooks on Information Systems. Springer Verlag, Heidelberg, 2004.

[11] Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. Sweetening Ontologies with DOLCE. In *Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, Proceedings of the 13th International Conference on Knowledge Acquisition, Modeling and Management* (EKAW 2002), LNCS-2473, Siguenza, Spain, 2002.

[12] ISO/IEC 15938-3 FCD Information Technology – Multimedia Content Description Interface - Part 3: Visual, March 2001, Singapore.