Advances in semantic multimedia analysis for personalised content access

Paola Hobson
Motorola, Centre for Applications, Content and Services
Motorola Labs , Basingstoke, UK
Paola.Hobson@motorola.com

Yiannis Kompatsiaris
Informatics and Telematics Institute
Thermi-Thessaloniki, Greece
ikom@iti.gr

Abstract— **In this special session overview paper, we present an introduction to the advances in semantic multimedia analysis for personalised content access. We firstly identify the importance of semantic multimedia analysis for personalised multimedia content applications, from a user and business perspective. We then summarise of the state of the art in these novel, and emerging domains. Finally, we describe some applications under development and evaluate their possible future commercial success.**

I. INTRODUCTION

User frustration in finding and organising multimedia content is one factor holding up advancement of markets for multimedia services, applications and equipment. Despite the huge quantity of multimedia content available via the internet and cable, satellite and broadcast services, users often cannot find what they would really like, and as a result, many valuable multimedia assets are under-utilised. Although users would benefit from personalisation, where a system matches available content to the user's stated and learned preferences, thereby enabling content offerings to be closely targeted to the user's wishes, such systems rely on the availability of metadata describing the content in order to make the match. Currently, metadata generation is mostly manual, which is costly and time consuming. Multimedia analysis techniques which go beyond the signal level approach to a semantic analysis have the potential to create automatic annotation of content, thereby opening up new applications which can unlock the commercial value of content archives.

Automated multimedia analysis tools, such as those described in the state of the art section below, are important enablers in making a wider range of information more accessible to intelligent search engines, real-time personalisation tools, and user-friendly content delivery systems. Such automated multimedia analysis tools, which add the semantic information to the content, are critical in realising the value of commercial assets e.g. sports, music and film clip services, where manual annotation of multimedia content would not be economically viable, and are also applicable to users' personal content (e.g. acquired from video camera or mobile phone) where the user does not have time to annotate all their content.

In this overview paper, we discuss the opportunities for enhanced personalised multimedia access and the benefits to both users and industry which can be achieved based on automated semantic analysis, and we summarise the technologies required in order to enable such applications. In the following section, the state of the art in relevant technologies is presented, followed by a brief summary of applications of personalised multimedia, using the aceMedia system [1] as an application example. Following this, success factors which are likely to influence the acceptance of potential future personalized multimedia applications are evaluated.

II. STATE OF THE ART

A. *Semantic Multimedia Analysis*

Despite intensive recent research, the automatic establishment of a correspondence between the low-level features and the semantic-level information (i.e. objects and events) needed to understand the content of the visual medium is a problem still far from being solved or adequately addressed [2], [3]. As a result of this inability to automatically bridge the semantic gap, some efforts have been focused on the introduction of the user in the loop, in combination with intermediate-level descriptors [4], or the restriction of the problem to the categorization of images in a small number of pre-defined classes, often in a hierarchical manner [5].

Most recent approaches try to exploit the user interaction or similar input of expert users by means of a learning approach [4]. Learning approaches rely on techniques such as neural networks [6], support vector machines [7], [8] hidden Markov models (HMMs) [9], genetic algorithms [10], and fuzzy systems [11]. The main characteristic of learning approaches is their ability to adjust their internal structure according to the given training set in order to approximate

This work was partially supported by the European Commission under contract FP6-001765 aceMedia.

the relations implied by the given training data. Consequently, they are suitable for use in problems such as object and event recognition, where the knowledge contained in any given training set is typically ill-defined or incomplete, i.e. where explicitly defining it is difficult.

Explicitly defined knowledge about objects and events of interest is generally not available, as indicated by the wide use of the aforementioned approaches, but this is true only for unrestricted domains. In well-structured domains such as sports, news, and personal content, the necessary knowledge can easily be assembled for the finite number of objects and events of interest. Ontologies can be used in these cases for explicitly expressing domain knowledge, facilitating the further inference of additional knowledge based on rules-based processing of pre-existing knowledge. In the current state of the art, however, ontology modelling and ontology-based metadata creation apply primarily to textual information [12], to which significant efforts have been devoted in the last few years, or are used only for assisting the manual annotation of photographs [13].

An understanding of the importance of generating a correspondence between domain-specific and low-level description vocabularies applied to multimedia, as well as the importance of exploiting a plurality of modalities for detecting cues useful to analysis, have recently revealed the possibility of using ontologies to drive the extraction of semantic descriptions. In [14], an object ontology, coupled with a relevance feedback mechanism to improve precision, is introduced to facilitate the mapping of low-level to high-level features and allow the definition of relationships between pieces of multimedia information. Recently, Semantic Web technologies have been identified as an appropriate framework for representing knowledge relevant to multimedia. In [15] an RDFS ontology for expressing MPEG-7 metadata terms is described, in order to make MPEG-7 accessible, re-usable and interoperable with other domains, while in [16] a methodology for enabling interoperability of OWL domain-specific ontologies with the complete MPEG-7 MDS is described. Extending this framework, basic mappings between domain-specific metadata vocabularies are provided for multimedia content in [17].

Despite the body of research described above, tangible applications of semantic multimedia analysis remain restricted. One reason is that methods which bridge the semantic gap are often tailored to a specific subject domain e.g. solutions specifically tailored to medical, sports or cultural domains, and are not generally applicable. There continues to be a strong need for research and development of domain independent methods. Another research problem is that multimedia analysis has traditionally taken a signal-level approach, using low-level image features as the basis of recognition methods, and knowledge engineering technologies such as those in development for the Semantic Web techniques are only now starting to be explored as promising areas in combination with low level analysis (e.g. [1])

B. *Personalisation for Multimedia Applications*

Personalisation is about developing models and systems to represent and capture user preferences, goals, needs, etc, and use this information in applications and services to better meet user needs and expectations, and help users achieve their tasks and goals more efficiently see e.g. [18].

With respect to multimedia content applications, personalisation can be applied to content filtering to reduce the amount of content proposed to a user [19], presentation of the information, content recommendation [20] and content search [21]. Hybrid recommender/filtering systems are also proposed [22][23] in which the benefits of reduction in information offered to the user are combined with the ability to recommend new content which the user might not have found if their filtering criteria are too tight. Personalisation systems generally rely on input and feedback from the user which enables their goals, preferences and interests to be acquired and reasoned on. Feedback can be explicit e.g. [27], where a user evaluates a content item presented to them, or implicit, where user actions are tracked in order to learn what was liked or intended. Much recent literature has focused on implicit profile adaptation e.g. [25]. However, [26] suggests that the use of implicit feedback may be limited in its scope. Generally the most successful system use a mixture of both techniques.

III. Applications of personalised multimedia

Applications of personalised multimedia access may be considered in two main categories: personalised content filtering to assist users with information overload and task efficiency, and proactive personalised content recommendation to assist with discovery and enjoyment of content.

In the former category, we consider both professional users and consumers who are seeking multimedia content within a large repository or potentially vast databases accessible via the internet. Conventional searching methods return large quantities of candidate content, and the user is confronted with the task of manually inspecting each to see if it fits their model of the desired content for the task or activity in hand. This can be exacerbated where the user can only express their query in imprecise terms e.g. an advertising executive looking for an "exciting" video of "water sports" or ambiguous terms e.g. pictures of a bank (financial institution or river feature?).

In a personalised system, knowledge about the user's interests, preferences, and past content consumption can be used to assist with a more intelligent search. Two main techniques can be offered. In the first, the system matches the user profile against the descriptions of the available content, using rule-based or other reasoning. Content which most closely matches the user's expressed or learnt desires is then selected. Such personalisation can work very effectively for established preferences and those which are easy to express, but cannot be used when the user is searching for new content which is not referred to in their

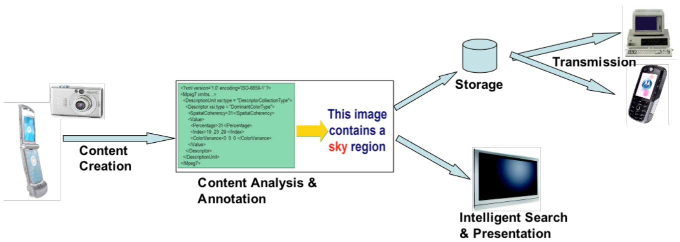

Figure 1. Overview of the aceMedia system.

profile or history. In such cases, more valuable personalisation can be provided by relevance feedback methods. The user is presented with a set of candidate content related to their query, and by providing positive and negative ratings, the system will iterate the query to return content closer to the positively rated items, and further from the negatively rated items.

A second main application area is associated with recommender systems where a user is recommended content according to what they have selected before or what others with similar tastes have selected e.g. [24]. Recommender type systems rely on a history of user interests being created over a period of time so that patterns of preference can be learned. Profiles must also be shared such that the system can compare one set of users with another in order to recommend new content e.g. "other people who liked movie A also liked movie B". The most well known recommender systems are Amazon and TiVo. Initialisation of these systems can be problematic, as there is no user history. Stereotypes can be used in such cases, where basic demographic data is used to establish a base set of preferences which will be adapted via user interaction over time.

Generally, personalisation systems can only be effective if the content has sufficient and meaningful metadata such that matching to user preferences, or reliable clustering in a recommender system, can be achieved. Therefore, long-term success of personalised systems is highly dependent on automated semantic metadata generation, which provides the basis of selection and recommendation.

The aceMedia project [1] addresses these issues by integration of Semantic Web and multimedia technologies in order to provide an efficient framework for semantic analysis and personalization, Figure 1. aceMedia develops tools to automatically analyze content, generate semantic metadata and annotation, and support for personalized and intelligent content search and retrieval services, Figure 2.

In aceMedia, the Knowledge Assisted Analysis module (KAA) creates automatic annotations using an ontology driven approach. Low level image features are extracted from the multimedia content, using tools such as segmentation and MPEG-7 descriptors extraction. Conversion of the MPEG-7 descriptors into an RDF representation enables reasoning to be applied such that objects and areas in the scene can be identified, with

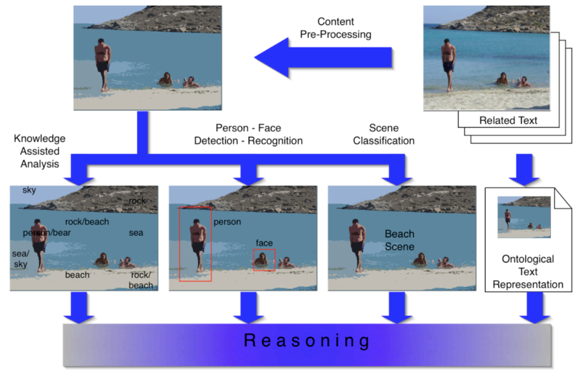

Figure 2 : aceMedia knowledge assisted analysis

reference to the appropriate domain ontology. More specifically, during image/video analysis, a set of atom-regions is generated by an initial segmentation while visual descriptors and spatial relations are extracted for each region. A distance measure between these descriptors and the ones of the prototype instances included in the domain ontology is estimated using a neural network approach for distance weighting. Finally, a genetic algorithm decides the labeling of the atom regions with a set of hypotheses, where each hypothesis represents a concept from the domain ontology. This approach is generic and applicable to any domain as long as new domain ontologies are designed and made available.

Within aceMedia, the automatically generated metadata is exploited by the personalisation module which creates a model of user preferences and profiles enabling personalised search and presentation of content. The user model is dynamically updated by learning on user behaviour as they interact with their content. Furthermore, semantic multimedia annotation in aceMedia is exploited in user centred applications such as intelligent search and retrieval, and automated content collections. aceMedia tools under development include user query interpretation, hybrid visual-semantic search and retrieval, and improved relevance feedback.

IV. SUCCESS FACTORS INFLUENCING PERSONALISED MULTIMEDIA APPLICATIONS

From the user perspective, there are two major obstacles to be overcome in provision of personalised multimedia applications. The first is a deep understanding of user preferences and the expression of those preferences by the end user. This must be accompanied by appropriate user modelling such that dynamic interpretation of user actions via reasoning technologies can lead to adaptation of the user profile in a meaningful way. It is very important to be able to recognise transitory changes in a user's interests compared to deep-rooted or long-term preferences Users may not have time to spend editing their profile, but it requires a long usage history to fully learn a user's preferences based on their actions.

The second issue concerns privacy and security of the user data. Personalising content selection, delivery and presentation relies on knowing something about the user. The user may be prepared to share some preference information with a service provider, but may be suspicious of a system which tracks their interactions with the service and automatically modifies their profile according to how these actions are interpreted. Systems which do not keep centralised histories e.g. where a user profile resides on the client device only, or those which only modify content searches based on the current session without retaining a permanent history at all, e.g. as a relevance feedback system may provide, are likely to have earlier user acceptance.

V. CONCLUSIONS

In this paper, we have skimmed the surface of semantic multimedia analysis for personalised content access, by reviewing some of the technical and user requirements, and by a summary of the state of the art in this emerging and fast growing area. We have presented a brief summary of the aceMedia system Knowledge Assisted Analysis in order to illustrate some emerging technical solutions.

Future research in this area includes the exploitation of reasoning techniques for multimedia applications. Therefore research will be carried out into uncertain and soft constraint reasoning using different techniques or combinations thereof. Whereas the current work focuses on annotating the detected segments with semantic labels, future research will additionally aim to enhance and correct the initial segmentations by merging the segments when appropriate based on spatial knowledge. Complementarily, DL reasoning can be used to check the consistency of the labels and to perform automatic classification of the annotated images. Finally, the use of representative pre-annotated cases coupled with adaptive rule adjustment in order to augment the knowledge that is already stored in the ontologies will be explored.

REFERENCES

[1] Integrating knowledge, semantics, and content, for user centred intelligent media services : the aceMedia project http://www.acemedia.org

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

[3] M.R. Naphade and T.S. Huang, “Extracting semantics from audio-visual content: the final frontier in multimedia retrieval”, IEEE Trans. on Neural Networks, vol. 13, no. 4, pp. 793-810, July 2002

[4] V. Mezaris, I. Kompatsiaris, N. V. Boulgouris, and M. G. Strintzis, "Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval", IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 5, pp. 606-621, May 2004.

[5] A. Vailaya, M.A. Figueiredo, A.K. Jain, and H.J. Zhang, "Image Classification for Content-Based Indexing", IEEE Transactions on Image Processing, 10(1):117–130, January 2001.

[6] R.D. Reed and R.J. Marks, Neural Smithing - Supervised learning in feedforward artificial neural networks, The MIT Press, 1998.

[7] V.Vapnik, “Statistical Learning Theory”, Springer-Verlag, New York 1998

[8] G.-D. Guo, A.K. Jain, W.-Y. Ma, and H.-J. Zhang, Learning similarity measure for natural image retrieval with relevance feedback, IEEE Trans. on Neural Networks, vol. 13, no. 4, pp. 811-820, July 2002.

[9] Rabiner, L., Juang, B., “An introduction to hidden Markov models”, IEEE ASSP vol. 3, no. 1, pp. 4-16, Jan. 1986.

[10] M.Mitchell, “An introduction to Genetic Algorithms”, MIT Press, 1996.

[11] J.C. Bezdek, J.M. Keller, R. Krishnapuram and N.R. Pal, “Fuzzy Models and Algorithms for Pattern Recognition and Image Processing”, Kluwer, 1999.

[12] S. Staab, R. Studer, H.-P. Schnurr and Y. Sure, “Knowledge processes and ontologies”, IEEE Intelligent Systems, vol. 16, no. 1, pp. 26-34, Jan-Feb 2001.

[13] A.T. Schreiber, B. Dubbeldam, J. Wielemaker and B. Wielinga, “Ontology-based photo annotation, IEEE Intelligent Systems”, vol. 16, no. 3, pp. 66-74, May-June 2001.

[14] Kompatsiaris, I., Mezaris, V. and Strintzis, M.G., 2004. “Multimedia content indexing and retrieval using an object ontology”, Multimedia Content and Semantic Web Methods, Standards and Tools, Editor G. Stamou, Wiley, New York.

[15] Hunter, J., “Adding Multimedia to the Semantic Web – Building an MPEG-7 Ontology”, In International Semantic Web Working Symposium (SWWS), Stanford, July 2001.

[16] Tsinaraki, C., Polydoros, P., Christodoulakis, S., “Interoperability support for Ontology-based Video Retrieval Applications”, In Proceedings of Third International Conference on Image and Video Retrieval (CIVR), Dublin, Ireland, July 21-23, pp 582-591, 2004.

[17] Hunter, J., “Enhancing the Semantic Interoperability of Multimedia through a Core Ontology”, In IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Conceptual and Dynamical Aspects of Multimedia Content Description. February 2003.

[18] Hirsh H., Basu C., Davison B. D., *Learning to Personalize,* Communications of the ACM, Volume 43, Issue 8, 2000.

[19] F. Liu, C. Yu and W. Meng. Personalized Web Search For Improving Retrieval Effectiveness. IEEE Transactions on Knowledge and Data Engineering, Vol. 16, N. 1, January 2004, pp. 28-40.

[20] A. Smeaton and J. Callan. Proceedings of the 2nd DELOS Network of Excellence Workshop on Personalisation and Recommender Systems in Digital Libraries. ERCIM Workshop Proceedings - No. 01/W03, European Research Consortium on Informatics and Mathematics, Dublin, Ireland, June 2001. http://www.ercim.org/publication/ws-proceedings/DelNoe02/

[21] G. Jeh, J. Widom. Scaling Personalized Web Search. Proc. of International World Wide Web Conference (WWW2003). ACM. Link-based ranking 1. Budapest, Hungary, May 20-24, 2003, pp. 271-279

[22] Baudisch P., *Joining Collaborative and Content-based Filtering*, Proceedings of Interacting with Recommender Systems, CHI '99 Workshop.

[23] Yu K., Schwaighofer A., Tresp V., Ma W.-Y., Zhang H., *Collaborative Ensemble Learning: Combining Collaborative and Content-based Information Filtering via Hierarchical Bayes*, Proceedings of UAI 2003, Morgan Kaufmann.

[24] Yahoo! Movies, http://movies.yahoo.com

[25] Jinmook Kim, Douglas W. Oard and Kathleen Romanik, Using Implicit Feedback for User Modelling in Internet and Intranet Searching, (2000).

[26] Diane Kelly, Jaime Teevan, Implicit Feedback for Inferring User Preference: A Bibliography

[27] Manos Papagelis & Dimitris Plexousakis, Recommendation Based Discovery of Dynamic Virtual Communities, http://www.csd.uoc.gr/~papaggel/docs/papers/all/CAiSE03-Recommendation_Based_Discovery_of_Dynamic_Virtual_Communities.pdf