7th WSEAS Int. Conf. on MATHEMATICAL METHODS and COMPUTATIONAL TECHNIQUES IN ELECTRICAL ENGINEERING, Sofia, 27-29/10/05 (pp330-335)

Knowledge Discovery in Multimedia Repositories: the role of metadata

B. VASSILIADIS[1], A. STEFANI[1], L. DROSSOS[2], K. IOANNOU[3]
[1]Computer Science
Hellenic Open University,
13-15 Tsamadou st. GR- 26222, Patras
GREECE

[2] Department of Applied Informatics in Management & Economy
Technical Education Institute of Messologi
GREECE

[3]Wireless Telecommunication Laboratory, Department of Electrical and Computer Engineering,
University of Patras, Rion 26500, Patras
GREECE

Abstract: - Data mining has recently emerged as an important issue of multimedia research. New trends are focused on knowledge discovery in large multimedia repositories, however, traditional data mining approaches do not seem to bear fruits. There is a need to use new techniques to mine knowledge from multimedia, link content top context and more importantly, to exploit them in real life applications. In this communication we review the state of the art in multimedia data mining, discuss the role of metadata standards and present future trends in this research area.

Key-Words: - Multimedia Mining, Knowledge discovery, Metadata, State of the Art.

1 Introduction

According to recent studies, humanity has produced, in the last three years, as much information as in all the previous years of its history. This totals an amount of 12 Exabytes of data in the form of print, sound and audiovisual media. The ever-increasing production of digital information, combined with the continuous improvement of digitization methods, has made the management of digital collections extremely difficult.

Multimedia, and more particularly images and video, are important parts of on-line digital information. Furthermore, advances in image acquisition and storage technology have led to tremendous growth in very large and detailed image databases. These digital media, if analyzed, can reveal useful information to the human users. Multimedia Mining (MM) deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the multimedia [1]. It is more than just an extension of data mining to the image or video domain. It is an interdisciplinary endeavour that draws upon expertise in image processing, image retrieval, semantic technologies and machine learning.

Multimedia mining or knowledge discovery is the nontrivial extraction of implicit, previously unknown, and potentially useful information from large collection of data. It can be viewed as a multidisciplinary activity because it exploits several research disciplines of artificial intelligence such as machine learning, pattern recognition, expert systems, knowledge acquisition, as well as mathematical disciplines such as statistics, information theory and uncertain inference. Knowledge discovery refers to the overall process of extracting high-level knowledge from low-level data in the context of large databases.

The overall goal of current research efforts is to enable automatic application to application (A2A) interaction in large digital repositories by way of machine processable knowledge. Research is also be targeted on the ability of the methodology envisaged to enable the efficient integration of discovered knowledge back into real life systems where it can be utilized. For this reason, a special case study will take place in the domain of digital art.

We view this research effort on multimedia data mining as more than just an extension of data mining to the multimedia domain. It is an interdisciplinary endeavor that draws upon expertise in image processing, image retrieval, semantic/metadata technologies and machine learning.

Data mining has recently emerged as an important issue of multimedia research. New trends are focused on knowledge discovery in large multimedia repositories, however, traditional data mining approaches do not seem to bear fruits. There is a need to use new techniques to mine knowledge from multimedia, link content top context and more importantly, to exploit them in real life applications. In this communication we review the state of the art in multimedia data mining and discuss future trends in this research area.

2 The different steps of Knowledge Discovery

Knowledge discovery process usually consists of an iterative sequence of, generally, the following steps (figure 1): data pre-processing, feature extraction, data mining tasks algorithms, interpretation and evaluation.

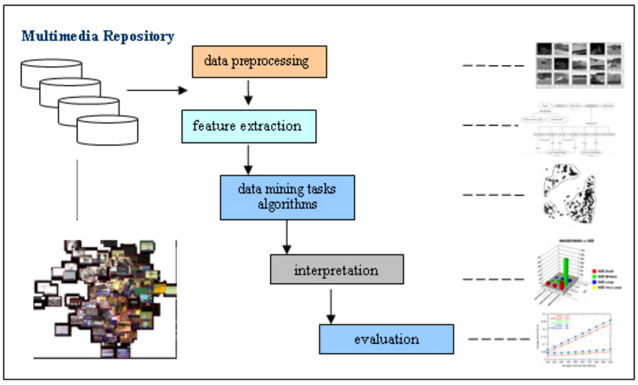

Figure 1. General multimedia mining process

Multimedia data mining, being a subfield of data mining, deals with the extraction of implicit knowledge, multimedia data relationships, or other patterns not explicitly stored in multimedia databases [2]. Multimedia data mining may not be limited to images, video or sound, but encompasses text as well. In this proposal, we consider only images and video objects when referring to multimedia data mining.

Despite the fact that multimedia has been the major focus for many researchers around the world, data mining from multimedia databases is still in its infancy. Although, it uses findings from the above mentioned individual scientific fields which, in themselves may be quite matured, multimedia mining, to date, is just a growing research focus and is still at an experimental stage. This has caused many misconceptions about what multimedia mining is and what are the differences with related areas. Works such as the one of [3] successfully point out major differences:

- MM vs. Computer Vision or Image Processing: MM's goal is to extract patterns from large collections while the other areas focus on understanding and/or extracting features from a single image.
- MM vs. Pattern Recognition: they share feature extraction steps but differ in pattern specificity. Pattern recognition's objective is to recognize specific, classification patterns; pattern generation and analysis. MM's objective is to generate all significant patterns without prior knowledge of what they are; patterns are more diverse; more aspects (besides generation and analysis) are involved: indexing, storage, representation, visualisation etc.

3 State of the Art

The diversity in the techniques used for MM has led to a plethora of methods that can, however, fall under two general frameworks: functional and information-driven. The former focuses on the functionalities of different component modules to organize image mining systems while the latter is designed as a hierarchical structure with special emphasis on the information needs at various levels in the hierarchy [4]. Since functional frameworks seem to be unable to take into account different layer of information representation, information driven frameworks have gained more acceptance.

Besides frameworks, several techniques have been used for every single step of the MM process. Since in this proposal we are also focusing on image processing techniques used in MM, the review will include object recognition, indexing and retrieval, image classification and clustering and association rules mining solutions proposed by researchers in the past few years. Although object recognition, indexing and retrieval are not directly considered by many as data mining tasks, we consider them as essential technologies for multimedia repository management. The combined use of these technologies with core data mining techniques leads to more efficient management processes.

Image indexing and retrieval solutions currently are used to solve similarity search problems. They are mostly feature-based solutions. The basic idea of feature-based similarity search is to extract important features from the multimedia objects, map the features into high-dimensional feature vectors, and search the repository of feature vectors for objects with similar vectors (figure 2).

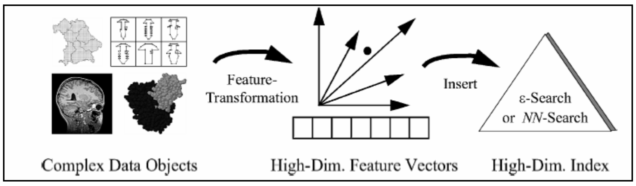

Figure 2. Feature-based similarity search ([7])

Features (figure 3) may include colour, histograms, edge distributions, texture measures and others. A good work on using numerous features at the same time for multimedia retrieval can be found in [5,6].

Figure 3. Decomposition of a multimedia object into features ([5])

K-dimensional data structures that support point, range and nearest neighbour queries are used to index the multimedia: SR-Tree, TV-tree, X-tree, Static Interpolation Search Trees etc. A comprehensive survey of data structures used in multimedia are presented in [7]. In the case of multimedia, feature similarity queries defines a metric space, instead of vector space. An interesting categorisation of different approaches, with a special interest in multimedia mining, are presented in [8]. One of the main problems in this area is the 'dimensionality curse', the efficient selection of features and of course retrieval speed and accuracy.

Recent approaches have explored the compressed domain, MPEG-7 based searching [9], multi-visual features [10], Wavelets and Discrete Fourier Transform [11,12], multi-resolution indexing, shape matching [13], models for semantics-based retrieval [14]. An exhaustive survey of image retrieval applications can be found in [15].

Image classification and clustering are the supervised and unsupervised classification of multimedia objects into groups. Recent approaches take into account the presence of metadata like the one of [16], automatic annotation and retrieval based on a training sets [17] and top–down orthogonal semantic classification of image fragments [18].

Association rule mining is frequently used in multimedia mining to uncover interesting trends, patterns and rules in large datasets. There are two main approaches: mining from large collections of multimedia objects, and mining from a combined collection of images and associated alphanumeric data (e.g. hypertext, hypermedia). Recent works in the first category include viewpoint patterns mining (patterns that capture the invariant relationships of one object from the point of view of another object) [19], content based approaches [20] visual dictionaries (created for colour, texture, shape feature etc.) [21], neuro-fuzzy approaches for limited number of images [22], patterns that can be analysed with wavelets [23], Markov models that discovers affinity relations [24] and unsupervised discovery (mining) of patterns in spatial-temporal multimedia data [25]. The second category includes research based on segmented images with associated text [26], support vector machines [27]; and combination with hypertext mining [28].

4 Linking content and context: the role of metadata

Metadata standards play an important role for multimedia mining because they are used to describe content and context. As context we refer to the application environment or the application itself. Context is important because it enforces certain requirements not only on how to treat data but also on what to derive form data in order to achieve adequate quality of service. For example, multimedia data mining in web applications uses the same type of information with a local multimedia repository (e.g. a digital library), Nevertheless, the context differs significantly since web applications use linked hypermedia structures which in themselves contain useful information. This information should be used in conjunction to the primary knowledge discovered in the media itself t derive more complete knowledge. It is the opinion of many researchers that, in order to exploit the potential of multimedia mining in real life applications, techniques should be application tailored, where possible. Otherwise, results may not be entirely useful. So an 'one size fits all' in multimedia mining does not seem to work.

Standardization bodies continue to work on media standards in order to provide a common approach to enable interoperability, better quality and efficiency under specific constraints. In recent years there has been a wide proliferation of MM standards, the major part of which can be grouped as follows:

- Video coding: MPEG-1, MPEG-2, MPEG-4, AVI
- Image coding: JPEG, TIFF, BMP, GIF, JPEG2000
- MM Presentations: SMIL and MHEG
- Metadata: MPEG7, MPEG21, JPX

Multimedia mining uses metadata both as a primary source of information and a method of description. Annotated media, as for example an image that contains MPEG-7 descriptions in its code stream, provide two sources of information to mining algorithms: raw content (low level information) and metadata descriptors (high level information). On the other hand, metadata may be used to describe the associations between media artefacts of the same repository (e.g. image A describes the same concept with the Region of Interest X of image B). Metadata can also be used to describe metadata (meta-metadata).

Metadata standardization has followed two separated approaches that confuse many non-experts and most importantly the media industry: MPEG group's and W3C.

The MPEG group is a member of the ISO, responsible for the development of well known and successful standards such as MPEG coding (MPEG4, MP3). Its efforts towards the development of metadata for multimedia has resulted in MPEG-7 and MPEG21. MPEG7 is a complex and large standard for metadata descriptors of multimedia. It tries to encompass the needs of the three communities: Artificial Intelligence, Image Processing and Knowledge Management. The result is a wealth of descriptors for both low level (visual characteristics) and high level (e.g. keywords) characteristics for three types of media: image, video and audio. MPEG21, is an upcoming standard that aims to support the whole content delivery chain from content creation to consumption by a wide range of devices and through a plethora of networks. Some of the key elements used include digital item declaration, identification, description, content handling, intellectual property management, digital item rights management and others. Both standards are based on XML (or XML-like) descriptions and above that, on Description Definition Languages (DDL) and XML Schema (e.g. MPEG-7 DDL).

The W3C effort is the 'Semantic Web', which aims to make A2A interaction possible through metadata. XML, RDF, RDF(S) and ontologies are some of the technologies one concepts that will possibly make the Semantic Web a reality. The ontology concept is of great importance to Multimedia Mining. Formally, an ontology is an explicit and commonly agreed definition of objects and concepts within a domain. It enables communication/sharing of knowledge, control of terminology and finally, coding of information in its structure. Thus, ontologies can be used to describe relations between media artifacts within a multimedia repository, express context information and more importantly support interoperability with other applications. Currently, the most important language for expressing ontologies is OWL while there is a plethora of visual tools for writing ontologies. Somewhat similarly to MPEG's standards, the Semantic Web is based on XML/RDF. The schema language adopted by W3C is RDF Schema and OWL.

One would expect that since both efforts are based on XML, they are compatible or supplement each other. Although the general goals of W3C are the same with MPEG group's the approach is different. First of all, W3C works on the Web context and does not pay so much attention to the content of the multimedia itself. For example, low level, visual feature descriptions are not explicitly taken into account in the sense that there are no explicit descriptors for them as in MPEG-7. Another obstacle is that during that the conversion from the schema language to XML is 'lossy'. This means that the reverse conversion (from XML to the Schema language) is possible but the description may differ significantly. Parsing is also difficult. Metadata descriptions in XML derived from MPEG-7 DDL may not be parsable by Semantic Web tools. MPEG's effort is more concentrated in the digital media domain, and as such it can be considered as a subset of the Semantic Web effort, although this is not entirely true. For example, MPEG-7 can be considered as an ontology and an ontology language at the same time.

Although these problems are of general interest, there is an impact on multimedia data mining. Decisions on what is the appropriate method for describing metadata and what are the implications in the application parameters have to be made. The continuing lack of standardization is prohibiting researchers to demonstrate the potential of new technologies. This also holds for multimedia data mining on the Web, a possible killer application for the future.

5 Conclusions

Multimedia Mining has naturally emerged as a separate research theme of Data Mining, the need to derive more information and knowledge from data. As our applications are more multimedia rich, the importance of multimedia mining has increased and the pressure for real life applications is demanding. The corresponding community relatively counts its first years of existence, however some major steps have taken place. The most important think is that researchers have realized that the road to multimedia mining requires strong collaboration and most of all standardization. These steps are necessary to move from the laboratory to commercial applications.

In this communication we provided a short presentation on Multimedia Mining, discussing its steps and its differences with other scientific disciplines. Special attention was given to metadata as a means for both deriving and describing knowledge in multimedia repositories. We argue that the different standardization efforts need to converge in the future in a more simple standard. On the other hand, application (context-specific) use of the standards has been seen by the industry as a more agile solution towards real applicability.

Acknowledgements

This work was supported by the EPEAEKII Programme "Programme for the support of the Informatics curriculum at the Department of Applied Informatics in Management & Economy, TEI Mesologiou" (ΣAE 445/3 code 2003ΣE44530107).

References:

[1] Han, J. and M. Kamber, (2001). Data Mining: Concepts and Techniques, Morgan Kaufmann.

[2] Zaïane, O.R., Han, J., Li, Z.-N., Hou, J., (1998). Mining Multimedia Data. Proceedings CASCON'98: Meeting of Minds, Toronto, Canada, pp. 83-96.

[3] Hsu, W., Lee, M.L., Zhang, J., (2002). Image Mining: Trends and Developments. Journal of Intelligent Information Systems, Vol.19(1), pp. 7–23.

[4] Zhang, J., Hsu, W., and Lee, M.L. (2001). An Information-Driven Framework for Image Mining. In 12th Int.Conference on Database and Expert Systems Applications.

[5] Tieu, K., Viola P., (2004). Boosting Image Retrieval. International Journal of Computer Vision, Vol. 56(1/2), pp. 17–36.

[6] Rui, Y., Huang, T., and Chang, S. 1999. Image retrieval: Current techniques, promising directions and open issues. Journal of Visual Communication and Image Representation, Vol. 10, pp. 39–62.

[7] Bohm, Ch., Berchtold, S., Keim, D.A., (2001). Searching in High-Dimensional Spaces—Index Structures for Improving the Performance of Multimedia Databases. ACM Computing Surveys, Vol.3(3), September 2001, pp. 322–373.

[8] Chavez, E., Navarro, G., Baeza-Yates, R., Marroquin, J.L., (2001). Searching in Metric Spaces, ACM Computing Surveys, Vol. 33(3), September 2001, pp. 273–321.

[9] Agius, H., Angelides, M.C., (2004), Modelling and Filtering of MPEG-7-Compliant Meta-Data for Digital Video. SAC '04, March 14-17, 2004, Nicosia, Cyprus, pp. 1248-1252.

[10] Ngu, A.H.H., Sheng, Q.Z., Huynh, D.Q., Lei, R., (2001). Combining multi-visual features for efficient indexing in a large image database. The VLDB Journal, Vol.9, pp. 279–293.

[11] Sabharwal , C.L., Subramanya, S.R., (2001). Indexing Image Databases Using Wavelet and Discrete Fourier Transform. SAC 2001, Las Vegas, NV, pp. 434-439.

[12] Zhang, D., Lu, G., (2005). Study and evaluation of different Fourier methods. Image and Vision Computing, Vol.23, pp. 33–49.

[13] Andreou, I., Sgouros, N.M., (2005). Computing, explaining and visualizing shape similarity in content-based image retrieval. Information Processing and Management 41 (2005), pp. 1121–1139.

[14] Meghini, C., Sebastiani, F., Straccia, U., (2001). A Model Of Multimedia Information Retrieval. Journal of the ACM, Vol. 48, No. 5, September 2001, pp. 909–970.

[15] Veltkamp, R.C., and M.Tanase, (2001). Content- Based Image Retrieval Systems: A Survey. TR UU-CS-2000-34, Utrecht Univ.

[16] Wang, L., Liu, L., Khan, L., (2004). Automatic image annotation and retrieval using subspace clustering algorithm. Proceedings of the 2nd ACM international workshop on Multimedia databases, pp. 100 – 108.

[17] Jeon, J., Lavrenko, V., Manmatha, R., (2003). Automatic image annotation and retrieval using cross-media relevance models. Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, July 28-August 01, pp. 119 – 126.

[18] Zhuge, H., (2004). Retrieve images by understanding semantic links and clustering image fragments. The Journal of Systems and Software, Vol.73, pp. 455–466.

[19] Hsu, W., Dai, J., Lee, M.L., (2003). Mining viewpoint patterns in image databases. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 553 – 558.

[20] Ordonez, C., Omiecinski, E., (1999). Discovering Association Rules Based on Image Content, Proceedings of the IEEE Forum on Research and Technology Advances in Digital Libraries, p.38.

[21] Zhang, R., Zhang, Z., Khanzode, S., (2004). A data mining approach to modelling relationships among categories in image collection. Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 749 – 754.

[22] Woodford, B.J., Deng, D., Benwell, G.L., (2004). A wavelet-based neuro-fuzzy system for data mining small image sets. Proceedings of the second workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalisation – Vol. 32, Dunedin, New Zealand, pp. 139 – 143.

[23] Li, T., Li, Q., Zhu, S., Ogihara, M., (2002). A survey on wavelet applications in data mining. ACM SIGKDD Explorations Newsletter, Vol.4(2), pp. 49 - 68.

[24] Shyu, M.L., Chen, S.C., Chen, M., Zhang, C., (2004). Affinity relation discovery in image database clustering and content-based retrieval. Proceedings of the 12th annual ACM international conference on Multimedia, pp. 372 - 375.

[25] Chang, S.F., (2003). Mining spatio-temporal patterns and knowledge structures in multimedia collection. Proceedings of the 1st ACM international workshop on Multimedia databases, pp. 1-1.

[26] Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D.M., Jordan, M.I. (2003). Matching words and pictures. The Journal of Machine Learning Research, Vol. 3, Pages: 1107 – 1135.

[27] Hoi, C.H., Lyu, M.R., (2004). Web image learning for searching semantic concepts in image databases. Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, pp. 406 – 407.

[28] Brown, R., Pham, B., (2005). Image Mining and Retrieval Using Hierarchical Support Vector Machines. Proceedings of the 11th International Multimedia Modelling Conference (MMM'05), pp. 446 - 451.