Caliph & Emir: Semantics in Multimedia Retrieval and Annotation

Mathias Lux
(Know-Center Graz, Austria
mlux@know-center.at)

Michael Granitzer
(Know-Center Graz, Austria
mgrani@know-center.at)

Werner Klieber
(Know-Center Graz, Austria
wklieber@know-center.at)

Abstract: As shown by recent studies it is estimated that there will be nearly 300 million digital image capture devices in use worldwide through 2004, capturing about 29 billion digital pictures, most of which will be organised in some kind of multimedia repository and available via the World Wide Web or through some other means of sharing data [Infotrend, 04]. Many of these images need to be stored, sorted and retrieved by using personal digital library applications. MPEG-7 offers a range of standardized description methods for generating metadata for multimedia content and allows ontology-like semantic descriptions. *Caliph & Emir*[1] is a pair of applications that use MPEG-7 for annotation and search of digital photos focusing on semantic descriptions.

1 Introduction

Digital camera users produce lot of images throughout a year and save them to personal computers. After some time the amount of photos exceeds the critical mass manageable without specialized tools. Most people create an intuitive structure for storing their personal image library. They create folders for images that are taken in the same context, for example "Photos from the I-Know '04 conference", however this does not enable the user to find a photo which shows a specific person, object or even expresses a specific idea or feeling when needed. Some file formats like TIFF and JPEG permit the user to enrich the visual information with textual descriptions, but they only offer limited capabilities. MPEG-7 offers a whole range of descriptors to annotate images with manually or automatically generated metadata. The picture can be described in many different ways regarding for example its quality, its technical attributes, its instances (thumbnails, high resolution, and so on) and its content from either a technical or a semantic point of view. Caliph, which is short for *Common And Light weight PHoto annotation*, allows to annotate digital photos manually and extracts content based on low level features from the image automatically. Emir, which is short for *Experimental Metadata-based Image Retrieval*, allows the retrieval of digital photos based on annotations created with Caliph.

[1] Caliph & Emir are available licensed under GPL: http://sourceforge.net/projects/caliph-emir

2 Related Work

There are many applications and projects similar to Caliph & Emir: Commercial Products like ACDSee[2], Google Picasa[3] and Adobe Photoshop Album[4] allow the annotation of photos using keywords or self-defined tags, which are visually attached to the image. For searching for these tags and keywords in an index, a retrieval mechanism is implemented. Digital Asset Management applications like the solutions of Blue Order[5] or Artesia[6] allow the storage and retrieval of digital images for productive environments based on metadata and keywords. Some companies like Convera[7] with its multimedia search and IBM with QBIC[8] implement content based image retrieval and image recognition for selected domains.

Another research project called VizIR studies the content based image retrieval capabilities of MPEG-7 (see e.g. [Eidenberger 04]). Mecca, which is a research project for the support of collaborative discourses in a scientific community (see [Klamma 03]), uses MPEG-7 to annotate videos based on a growing and evolving ontology for collaborative knowledge creation. Although they deal with semantic knowledge about video assets the MPEG-7 descriptors for semantic annotations are not used. Most other research projects like the MPEG-7 Multimedia Data Cartridge (see [Kosch 04]) or the Intelligent Multimedia Database IMB (see [Klieber 04]) focus on video data. Besides the IMB project, which uses parts of Caliph for creating semantic annotations, none of the above mentioned projects allows the semantic annotation of multimedia content. A research project, which deals with MPEG-7 based semantic descriptions for interactive TV, is described in [Tsinaraki 03]. The authors introduce a framework for managing semantic descriptions based on a static domain using a fixed ontology including a data retrieval API for semantic descriptions. The visual creation of semantic descriptions, which do not underlie restrictions from a previously defined domain ontology, and information retrieval on semantic descriptions is not supported by the framework.

3 Caliph

Besides creating new annotations with Caliph, the requirements include that existing annotations should not be ignored. Caliph is able to extract existing EXIF[9] and IPTC IIM[10] annotations from images and converts them into valid MPEG-7 documents. Besides importing existing information the ColorLayout, the ScalableColor and the EdgeHistogram descriptors are extracted from the image. For manual annotation the author can fill in text fields for free text, structured descriptions of the image, and rate the image quality on a scale from 1 to 5. The core element of Caliph is the semantic annotation panel. On the right hand side of this panel the annotation author can create and maintain a library of reusable MPEG-7 based semantic objects. These objects can be dragged to the drawing panel, where they are positioned automatically using a

[2] http://www.acdsystems.com
[3] http://www.picasa.com
[4] http://www.adobe.com
[5] http://www.blue-order.com
[6] http://www.artesia.com
[7] http://www.convera.com
[8] http://wwwqbic.almaden.ibm.com
[9] http://www.exif.org
[10] http://www.iptc.org

modified force directed placement algorithm for graph embedding, which was originally specified in [Eades 84]. The semantic objects are visualized as graph nodes whereas the MPEG-7 relations between the nodes, which can be drawn by the user with the mouse, are visualized as edges, which results in a directed graph as visualization metaphor for MPEG-7 based semantic descriptions.

The spring embedding algorithm can be described as follows: In the first step the graph nodes attract or repel each other if and only if the nodes are connected through an edge, forming a virtual spring. Depending on their pair wise distances in the layout the virtual springs affect connected nodes by contraction or distraction with logarithmic strength. In the second step non adjacent nodes repel each other with inverse square force (see also [Eades 84]). According to these calculated force vectors each node moves along its force vector to create a new layout in a third step. Iterating these 3 steps a fixed number of times creates a force directed placement of the input graph.

The modifications include an improved, adaptive stop condition, which stops the computation when the layout stabilizes and an invisible attracting node in the centre of the drawing panel which affects all nodes with an invisible spring.

The new stop condition allows lets the algorithm itself decide how many iterations have to be taken to gain a force equilibrium. Using the attracting node in the centre results in a more circular layout for not connected graphs. Without this centre node the not connected sub graphs repel each other too strong for creating a visually appealing layout.

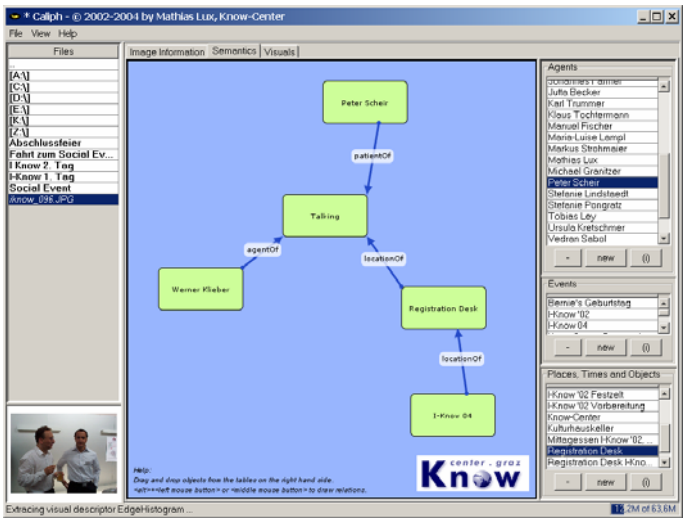

Figure 1 - Semantic annotation panel of Caliph with a graph layout generated by the spring embedding algorithm

As in MPEG-7 based semantic graphs not more than 10-20 nodes are used in average the cubic time complexity does not pose a problem.

A common problem with the annotation of a great number of documents is the time consumption of the manual annotation task. Caliph supports pre-annotation of sets of images in so far that a common description can be applied to a set of photos. The so-called autopilot extracts the metadata from the images and stores a predefined common description which is given by the user and can be applied all images of the set. For example: If all the photos where taken at the I-Know '04 conference the common description would set the field "Where" to *Graz*, *Austria* and the "When" field to *July*,

2004. In Addition the Semantic place *Graz* and the Semantic event *I-Know '04* can be added to the semantic description. After applying this common description to all photos of the set the user can open each of the already existing descriptions and add document-specific information. By opening a already existing MPEG-7 document instead of generating a new one, Caliph can take the already existing descriptors for ColorLayout, ScalableColor and EdgeHistogram instead of extracting them, which results in a major speed up of the annotation task.

4 Emir

A set of photo files annotated with Caliph can be easily searched by using Emir. Basically Emir allows an experimental retrieval of MPEG-7 descriptions based on either keywords, generic XPath statements, content based image descriptors and simple semantic graphs. Technically, for the sake of simplicity, the retrieval is disk based without an index. Keyword based queries, image content and semantic graphs are translated to XPath statements which are executed on each of the MPEG-7 files in one directory and its subdirectories.

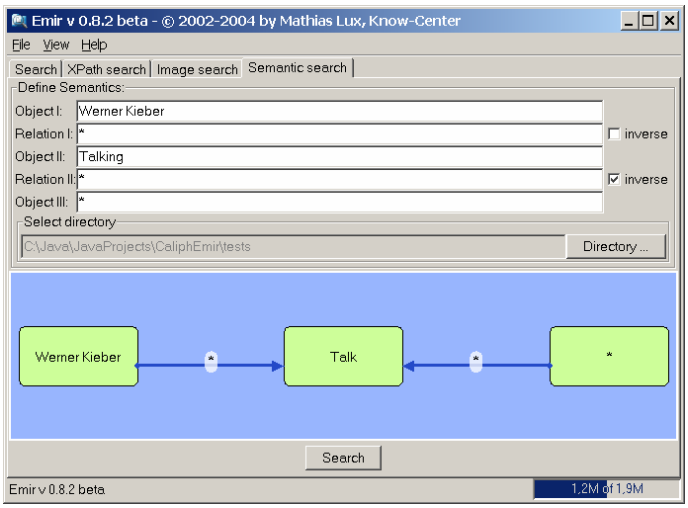

Figure 2 - Defining a query for a semantic description in Emir

The keyword query and the XPath query are simple use cases and can be implemented easily using existing retrieval mechanisms. The content based image query mechanism is specified in the MPEG-7 Standard (see [MPEG 2001a], [MPEG 2001b]) and can be implemented by given guidelines. Alternative approaches for similarity measurement for ColorLayout, ScalableColor and EdgeHistogram, and optimizations of the MPEG-7 based CBIR task are the subject of research in the VizIR project [Eidenberger 04]. The search for semantic descriptions lacks any guidelines by the MPEG consortium. Due to limited time resources only a trivial information retrieval mechanism and result ranking has been implemented. A maximum of 3 nodes and 2 edges is allowed for query formulation. Nevertheless the system shows that with given possibilities a formal definition of simple queries like "Give me all pictures where Person A is talking to someone" is possible and yields interesting results. An example for such a query is

“Give me all pictures where someone is talking” or “Give me all pictures where person A interacts with person B”.

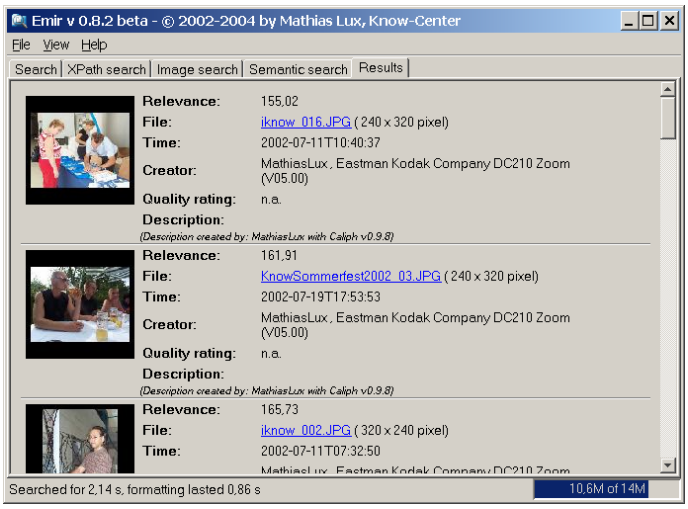

Figure 3 - Search result for a content based image query using the EdgeHistogram descriptor

Possible future steps include the usage of an index based search engine like Jakarta Lucene[11] for keyword based search to increase the retrieval performance. Another crucial point is the retrieval of semantic descriptions. To allow the retrieval of semantic graph structures an underlying retrieval model based on models such as the vector space model or the probabilistic model (see [Salton 75], [Rijsbergen 79]) will be introduced and evaluated. Furthermore additional media types will be supported by Caliph and Emir, a possible next step is the integration of audio and video files.

5 Outlook

Other projects at the Know Center the authors of this paper are working on include the clustering and visualization of cross media data sets. In our understanding the term cross media stands for an equal treatment of the essence, the actual meaning of the content, independent from the media used for deliverance. Based on the inter- and intra-modal similarity measures a set of documents, containing digital photos and HTML documents is clustered and a 2-dimensional visualization is calculated and presented.

[11] http://jakarta.apache.org/lucene

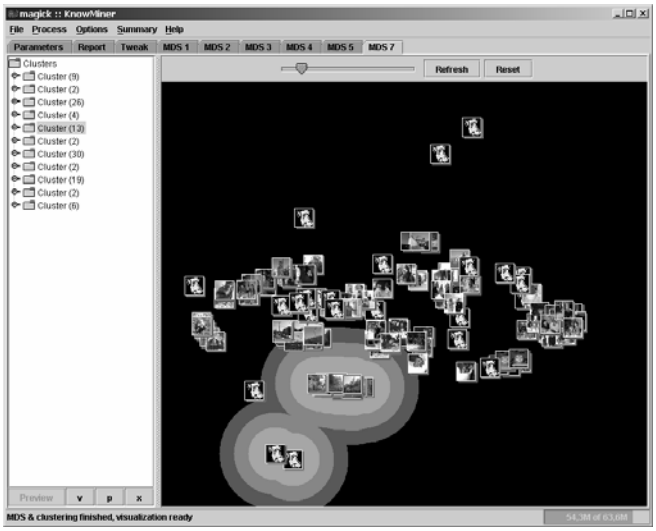

Figure 4 - The prototype Magick allows the clustering and visualization of cross media data sets

After defining a basic retrieval model for MPEG-7 documents visual, auditory and audiovisual documents with MPEG-7 descriptions can be clustered and visualized using common information retrieval techniques (see also [Lux 04]).

6 Acknowledgements

We would like to thank our colleagues at the Know-Center, and Graz University of Technology for their feedback and suggestions. We also want to thank Open Source Technology Group for providing SourceForge.net for hosting open source projects for free especially Caliph & Emir at http://sourceforge.net/projects/caliph-emir.

The Know-Center is a Competence Center funded within the Austrian K plus Competence Centers Program (www.kplus.at) under the auspices of the Austrian Ministry of Transport, Innovation and Technology.

References

[Eades 84] Eades, P., “A heuristic for graph drawing”, Congressus Nutnerantiunt, 42, 149–160, 1984.

[Eidenberger 04] Eidenberger, Horst, “Statistical analysis of MPEG-7 image descriptions”, ACM Multimedia Systems journal, Springer, 2/2004, URI: http://www.ims.tuwien.ac.at/media/documents/publications/acmms2004.pdf

[Hawkins 04] EXIF.org, Hawkins John, last visited 10.10.2004, http://www.exif.org/

[Infotrend 04] Infotrend Research Group, Inc., 2004, Internet, www.infotrends-rgi.com

[Kosch 04] Kosch, Harald, Döller, Mario, “An MPEG-7 Multimedia Data Cartridge”, Proceedings of the SPIE Conference on Multimedia Computing and Networking 2003 (MMCN03), Santa Clara, CA (USA), January 2003. SPIE Press.

[Klamma 03] Klamma, R., Spaniol, M., Jarke, M., “Digital media knowledge management with MPEG-7”, The Twelfth International World Wide Web

Conference, (WWW 2003), Poster Session, May 2003, Budapest, Hungary, URL: http://www-i5.informatik.rwth-aachen.de/lehrstuhl/publications

[Klieber 04] Klieber, Werner, Tochtermann, Klaus, Lux, Mathias, Mayer, Harald, Neuschmied Helmut, Haas, Werner, „IMB - Ein XML-basiertes Retrievalframework für digitales Audio und Video“, Berliner XML Tage 2003, Berlin, Germany

[Lux 04] Lux, Mathias, Granitzer, Michael, Sabol, Vedran, Klieber, Werner, Granitzer Michael, “Cross Media Retrieval in Knowledge Discovery”, Proceedings of the 5th International Conference on Practical Aspects of Knowledge Management, Vienna, Dec. 2005

[MPEG 2001a] MPEG Consortium, "ISO/IEC 15938: Information Technology - Multimedia Content Description Interface," 2001.

[MPEG 2001b] MPEG Consortium, "ISO/IEC 15938: Information Technology - Multimedia Content Description Interface - Part 2: Video," 2001.

[Rijsbergen 79] van Rijsbergen, C.J., “Information Retrieval”, London: Butterworths, 1979

[Salton 75] Salton, G., Wong, A. and Yang, S.S., 'A vector space model for automatic indexing', Communications of the ACM, 18, 613-620 (1975).

[Tsinaraki 03] Tsinaraki, Chrisa, Fatourou, Eleni, Christodoulakis, Stavros, “An Ontology Driven Framework for the Management of Semantic Metadata Describing AudioVisual Information”, in proceedings of the CAiSE 2003, LNCS 2681, Springer-Verlag, pp. 340-356, 2003