

On the Complexity of Annotation with the High Level Metadata

Mathias Lux

(Klagenfurt University, Austria
mathias.lux@itec.uni-klu.ac.at)

Werner Klieber

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

Michael Granitzer

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

Abstract: High level metadata provides a way to manage, organize and retrieve multimedia data based on the actual content using content descriptions. The MPEG-7 Semantic Description Scheme provides tools for storing expressive and interpretable high level metadata. As it is currently impossible for computers to create high level metadata autonomously, users have to create the annotations manually. Generally the manual annotation of multimedia content is understood as laborious and complex task. Within this publication we assess the complexity of the annotation task for the MPEG-7 Semantic Description Scheme within a small user evaluation and the results of the evaluation are discussed.

Keywords: MPEG-7, Metadata, Semantic Gap, Annotation, Evaluation

Categories: H.3.1, H.3.2, H.3.3, H.3.7, H.5.1

1 Introduction

While the MPEG consortium provided the formal structure for annotation of multimedia content with MPEG-7 (see [Chang 2001] for an overview on MPEG-7) as well as the possibility to adapt and enhance the formalism, the actual annotation task with high level metadata, which has to be done at least partially manually, is not addressed by the consortium. High level metadata ranges from classifications based on classification schemes, to textual content descriptions and further to ontological representations of concepts addressed in the content. Low level metadata on the other hand is based on features, which can be extracted from the content. This includes color histograms of frames and still images, rhythm or timbre of an audio stream or even terms extracted from text documents. While low level metadata can be extracted from video and audio streams, annotation with high level metadata cannot be done by the computer autonomously. Del Bimbo addresses this shortcoming within his book *Visual Information Retrieval* (see [DelBimbo 1999]) and defines the *semantic gap*. The semantic gap implicitly defines the fuzzy border between high and low level metadata and the transition from data to semantics and understanding of content.

Within this publication the time taken for annotation is assessed by a user evaluation. In the following chapter annotation tools for MPEG-7 based multimedia descriptions are introduced, followed by a description of one selected tool called *Caliph*, which was used for the evaluation. After a description of the evaluation methodology, results are presented and a conclusion is drawn.

1.1 Related Work

There are numerous annotation tools for the creation of MPEG-7 documents available. Some, like the IBM VideoAnnex (see [Lin 2003]) or MECCA (see [Spaniol 2005]), do not implement the MPEG-7 Semantic Description Scheme. A research project, which deals with MPEG-7 based semantic descriptions for interactive TV, is described in [Tsinaraki 2003]. The authors introduce a framework for managing semantic descriptions based on a static domain using a fixed ontology. They also include a data retrieval API for semantic descriptions. The visual creation of semantic descriptions without restrictions derived from previously defined domain ontologies is not supported by the framework. Within the Intelligent Multimedia Database IMB (see [Mayer 2003]), which focuses on video data, semantic annotation is supported through the integration of parts of Caliph, a “Common And Lightweight Interactive PHoto annotation tool (see [Lux 2003]). Caliph supports MPEG-7 based structured annotation of images (e.g. Photos, Video Key Frames, etc.) using graph structures and serves as basis for the evaluation presented in this paper.

To the best of the authors’ knowledge none of the above mentioned tools was investigated within a user evaluation. Therefore no reliable facts about the complexity of annotation using MPEG-7 are available. Our publication aims to fill this gap.

2 Caliph: The Annotation Tool

The annotation tool *Caliph* allows the creation of MPEG-7 descriptions for digital photos. Besides the ability to describe the content of the photos textually, an editor for semantic descriptions based on the MPEG-7 Semantic Description Scheme is integrated. The editor uses the metaphor of “drawing” a concept graph with semantic objects as nodes and semantic relations as arcs. Nodes can be re-used as they are stored in a node catalogue.

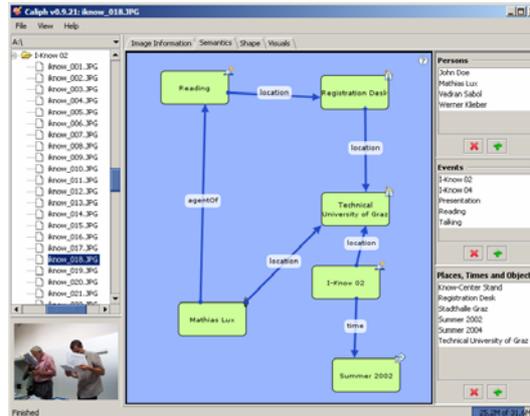


Figure 1: A screenshot showing the semantic description editor of Caliph.

As can be seen in above figure, the annotation editor panel is in the centre of the tool. On the right hand side the catalogue for semantic objects is shown, where users can add and delete objects. On the left hand side the image preview and the file navigation tree are shown. To annotate an image users have to open the file using the file navigation tree. After initial metadata extraction (MPEG-7 Descriptors Scalable Color, Color Layout and Edge Histogram as well as the EXIF and IPTC metadata encoded in the image file) users can edit the textual descriptions of the image in a first step. After rating the quality of the image on a subjective scale ranging from 1 (best) to 5 (worst), the semantic annotation takes place. Before the actual graph drawing can be done, the nodes required for the annotation task are added to the node catalogue and then dragged to the drawing panel. Using mouse interaction the relations between nodes can be drawn. For additional information see [Lux 2003].

3 Evaluation

To evaluate the medium time taken for annotating a digital photo, 5 users were given two annotation tasks and the time for fulfilling the 2nd task was measured. After a short written tutorial, which was read by each user, the first task was assigned. For this task the user was supported by an expert. The second task had to be completed without help. The two tasks goals were to annotate one image in each task. Each annotation includes a structured text annotation (with fields for “who”, “when”, “why”, etc. as defined in the MPEG-7 standard) and a free text annotation of the image contents, a subjective quality rating of the image, the name the metadata author and the image creator as well as above mentioned semantic annotation. Within a questionnaire the users gave feedback after the evaluation.

The results of the evaluation can be summarized easily: For the first task the users needed a medium time of 15.4 minutes, whereas for completing the second task a medium time of 6 minutes was needed. In contrast to these values an expert user needed around 30 minutes to annotate 17 photos, which results in ~ 1.7 minutes

annotation time. These findings tempt to draw the conclusion that there is a steep learning curve for semantic annotation with a minimum annotation time greater than 1 minute. Using the annotation time of the expert user, an average time for annotating thousand digital photos with small semantic graphs, which have a size smaller than 10 nodes, is ~ 28.3 hours. In other words an expert annotator could describe estimated 235 photos within one work day (calculated with 8 hours a day, having 50 minutes each).

Within the final questionnaire the users were asked rate following statements on a scale from 0 to 6 depending on if the users strongly agreed (0) with the statement or if they strongly disagreed (6) with the statement.

Statement	Rating A	Rating B
The complexity of semantic annotation is too high to be useful for organizing digital photos.	3.6	4.8
I would find it easy to annotate a large set of digital photos (e.g. 100+).	3.6	3.2
I would recommend Caliph or a similar tool to annotate digital photos.	2.2	1.6
I can see an obvious benefit by using semantic meta data for the organization of digital photos.	1.6	0.8

Table 1: Results from evaluation questionnaire

The statements were provided in two different contexts. With rating A users were asked to imagine the statements in context of managing their personal digital photos for their own use, while with rating B users were asked to imagine the use within a large company, where photos have to be organized and managed. As can be seen easily from the results in above table the interviewed users tend to see the use of the tool in a more professional setting.

4 Conclusion

The evaluation presented in this paper acknowledged that semantic annotation of digital photos is a laborious and longsome task. Within the first task users had to learn how to use the tool and had to understand the idea and concepts of the MPEG-7 Semantic Description Scheme. Although the users participating in the evaluation were satisfied with Caliph in general they pointed out a lot of features that were missing in their opinion. Examples for missing features were an editor for the object catalogue or an undo function for the drawing panel. The recommendations of the users show that an annotation tool for semantic metadata should be designed and implemented using user centred development (see [Holzinger 2005] for details), where user feedback is collected and integrated in early design and development stages and therefore ease-of-use is optimized.

Acknowledgements

The Know-Center is a Competence Center funded within the Austrian Competence Center program K plus under the auspices of the Austrian Ministry of Transport, Innovation and Technology (<http://www.ffg.at/index.php?cid=95>) and by the State of Styria.

References

- [Chang 2001] Chang, S.F., Sikora, T., Puri, A.: Overview of the MPEG-7 standard, Special Issue on MPEG-7, IEEE Transactions on Circuits and Systems for Video Technology, IEEE, pp. 688-695, 2001.
- [DelBimbo 1999] Del Bimbo, A.: Visual Information Retrieval, Morgan Kaufmann, 1999
- [Holzinger 2005] Holzinger, A.: Usability engineering methods for software developers, Communications of the ACM, ACM Press, 2005, Vol. 48, pp. 71-74
- [Lin 2003] Lin, C.-Y., Tseng, B. L., Smith, J. R.: VideoAnnEx: IBM MPEG-7 Annotation Tool for Multimedia Indexing and Concept Learning, in Proc. IEEE Intl. Conf. on Multimedia and Expo (ICME), July, 2003.
- [Lux 2003] Lux, M., Becker, J., Krottmaier, H.: Semantic Annotation and Retrieval of Digital Photos, in Proc. CAiSE 03 Forum Short Paper Proceedings Information Systems for a Connected Society, 2003, URI: <http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-74/>
- [Mayer 2004] Mayer, H., Bailer, W., Neuschmied, H., Haas, W., Lux, M., Klieber, W.: Content-based video retrieval and summarization using MPEG-7, in Proc. Internet Imaging V, SPIE 16th Annual Symposium, Electronic Imaging, USA, 2004
- [Spaniol 2005] Spaniol, M., Klamma, R., Waitz, T.: MECCA-Learn: A Community Based Collaborative Course Management System for Media-Rich Curricula in the Film Studies, in Proc. Advances in Web-Based Learning - ICWL 2005, 4th International Conference, Springer, Hong Kong, China, August 2005, LNCS 3583, pp. 131-143
- [Tsinaraki 2003] Tsinaraki, Chrisa, Fatourou, Eleni, Christodoulakis, Stavros: An Ontology-Driven Framework for the Management of Semantic Metadata Describing Audiovisual Information, Springer, LNCS, 2003