Journal of Universal Computer Science, vol. 14, no. 10 (2008), 1792-1807
submitted: 3/10/07, accepted: 11/2/08, appeared: 28/5/08 © *J.UCS*

Imagesemantics: User-Generated Metadata, Content Based Retrieval & Beyond

Marc Spaniol, Ralf Klamma
(Lehrstuhl Informatik 5, RWTH Aachen University, Germany
{spaniol|klamma}@i5.informatik.rwth-aachen.de)

Mathias Lux
(ITEC, Klagenfurt University, Austria
mlux@itec.uni-klu.ac.at)

Abstract: With the advent of Web 2.0 technologies a new attitude towards processing contents in the Internet has emerged. Nowadays it is a lot easier to create, share and retrieve multimedia contents on the Web. However, with the increasing amount in contents retrieval becomes more challenging and often leads to inadequate search results. One main reason is that image clustering and retrieval approaches usually stick either solely to the images' low-level features or their user-generated tags (high-level features). However, this is frequently inappropriate since the "real" semantics of an image can only be derived from the combination of low-level and high-level features. Consequently, we investigated a more holistic view on image semantics based on a system called *Imagesemantics*. This system combines MPEG-7 descriptions for low-level content-based retrieval features and MPEG-7 keywords by a machine learning approach producing joined OWL rules. The rule base is used in *Imagesemantics* to improve retrieval results.

Key Words: Web 2.0, Social Media Platform, User-Generated Content, MPEG-7
Category: H.3.3, H.3.4, H.3.5, H.5.1

1 Introduction

"What the heck do these images have to do with what I'm looking for?" That is a question many of us frequently ask themselvs when querying for images on the Web. Del Bimbo calls this the "semantic gap", the difference between technical extraction of data and the semantically correct interpretation of content [DelBimbo 1999]. Regardless of searching for pictures via Google Image Search [Google 2007], Yahoo! Photos [Yahoo 2007] or Flickr [Flickr 2007], the retrieval results have a low precision and thus are unsatisfying for most users. Missing or low quality metadata is the most common reason for a low precision in image retrieval. Many search engines for instance only employ the website, which links to the image, as single source of metadata. Even more, common search interfaces are mostly restricted to textual queries. Contrariwise, the open source projects LIRe and Caliph & Emir allow content based image retrieval given an image (dataset) [Lux et al. 2003, Lux et al. 2006]. The combination of both strategies (text and image analysis) is rather rare but has for example been applied in IBM's

Proventia Web Filter [IBM 2007]. For that purpose, the Proventia Web Filter is only suitable for "defensive" content blocking instead of "active" ad-hoc searches. Another approach for the combination of low-level and high-level metadata has been made in *Magick*, an application for cross media visual analysis (see also section 3.3).

An application for common retrieval tasks, which supports ad-hoc search not based on filtering, does currently not exist to the best of our knowledge. Therefore we developed *Imagesemantics*: A speedy and concise image retrieval system that allows searching for images in order to narrow (in order to bridge) the "semantic gap" between low-level content based features and high-level metadata annotations.

In this paper, we first give an overview on current state-of-the-art image retrieval techniques. Then, we introduce related image retrieval systems. After that, we present our *Imagesemantics* system, which incorporates OWL-based rules for the combination of high-level features (vocabulary independent keywords, called tags) and low-level image features. The paper closes with conclusions and gives an outlook on further research.

2 Image Retrieval Techniques Compared

In general two different types of image retrieval can be distinguished: (i) Retrieval based on content-dependent metadata and (ii) retrieval based on content-descriptive metadata [DelBimbo 1999]. Content-dependent metadata includes low-level features automatically generated from the image content. For content-dependent metadata no user interaction is needed. Typical examples are low level features like as an image's color feature vectors, where color characteristics of an image are expressed through a numerical vector. Content-descriptive metadata are typically manual annotations and have a high level of semantics. Examples are image description by text or through an ontology. In the following, we will introduce both concepts, with a focus on standard compliant information processing. In this aspect, we will stick to MPEG-7 because it offers the semantically richest metadata model for describing the content of audio-visual media.

2.1 Content-dependent Metadata: Low-level MPEG-7 Features

MPEG-7 (also called the *Multimedia Content Description Interface*) [ISO 2002, ISO 2003] is an international standard for storage and transmission of audio and visual content descriptions. It offeres an extensive metadata model covering a whole range of aspects (e.g. production, distribution, storage, rights, transmission, usage) and is the first standard from MPEG which considers multimedia metadata [Chang et al. 2001]. It provides a rich set of description schemata,

which allow to describe the content in structural (space and time) as well as semantic aspects. The MPEG-7 metadata model is based on *Descriptors* (*D*) which define the syntax and the semantics of feature representations and *Description Schemes* (*DS*) which specify the relationships between components (both *D* and *DS*). In addition, *Classification Schemes* (*CS*) allow simple creation of taxonomies to extend existing ones in MPEG-7. Due to the fact that new Descriptors, Description Schemes and Classification Schemes can be added easily, MPEG-7 is extensible to support arbitrary domains and use cases.

For the storage and processing of low-level, content-dependent metadata MPEG-7 basically provides three different types of features: **Color**, **texture**, and **shape** descriptors. **Color descriptors** provide a means to describe images based on their color characteristics. In this regard, MPEG-7 allows the standard compliant processing of seven distinct features:

- *Color Space* specifies in which format the color descriptors are expressed.
- *Color Quantization* defines the quantization of a color space.
- *Dominant Color* specifies a ranked set of dominant colors in an arbitrarily shaped region as well as a measure for the spatial coherency of the colors.
- *Scalable Color* defines a color histogram in the HSV color space.
- *Color Layout* describes the spatial distribution of colors for high-speed retrieval and browsing based on dominant colors in fixed regions.
- *Color Structure* specifies color content and the spatial arrangement of this content.
- *Group-of-Frame/Group-of-Picture* describes the color features of a collection of (similar) images or video frames by means of the scalable color descriptor.

Despite its name the *Group-of-Frame/Group-of-Picture* is not necessarily the most suitable descriptor for videos. The reason is that this descriptor is just a "simple" aggregation of the *Scalable Color* and is intended to generate descriptions for short video clips and animations instead of managing collections. In general, any of these descriptors have proved to work well on photos [Eidenberger 2004].

In addition, MPEG-7 offers three descriptors describing **texture** characteristics of an image:

- *Homogeneous Texture* characterizes the region texture using the energy and energy deviation in a set of frequency channels.
- *Texture Browsing* specifies the perceptual characterization of a texture.
- *Edge Histogram* specifies the spatial distribution of five types of edges in local image regions.

Here the *Edge Histogram* descriptor performs best, while *Homogeneous Texture* is highly redundant and *Texture Browsing* partially delivers ambiguous results which are more suitable for browsing instead of retrieval [Eidenberger 2004].

The third and last group of descriptors in MPEG-7 is used to describe **shape** characteristics of visual information:

- *Region Shape* specifies the region-based shape of an object.
- *Contour Shape* defines a closed contour of a 2D object or region.
- *Shape 3D* specifies an intrinsic shape description for 3D mesh models.

These descriptors only have limited use for image retrieval: Intuitively, the *Shape 3D* descriptor is not useful for image retrieval. Similarly, the *Contour Shape* descriptor cannot be employed for image retrieval easily as no transformation into data vectors and no distance measure for this descriptor is standardized in MPEG-7. Thus, only the *Region Shape* descriptor makes sense for image retrieval, but has proved to be highly dependent on the *Color Layout* descriptor.

As this reflects only the MPEG-7 perspective of content based image retrieval we recommend that the interested reader takes a look at [Smeulders et al. 2000], where content based image retrieval features and approaches are summarized.

2.2 Content-descriptive Metadata: High-level Annotations

In order to overcome the problems with interpreting semantics of audio-visual contents correctly high-level metadata in the form of annotations are used. While the extraction of low-level features can be done automatically, the annotation of images with high-level metadata annotations is mainly a manual procedure. High-level metadata annotations are a means of classifying, organizing and (finally) retrieving audio-visual contents. Similarly as with the content-based analysis of images by low-level features, the MPEG-7 standard provides dedicated descriptors for high-level metadata annotations. These annotations reach from textual content descriptions up to "ontologies". Thus, MPEG-7 offers a wide range for high-level interpretable and interoperable metadata annotations.

3 Image Retrieval Systems

Most of the existing image retrieval systems are based on either textual (metadata) descriptions or the image's low-level features. Image retrieval systems based on textual metadata in general employ multi field document retrieval [Robertson et al. 2004]. Multiple fields denoted by the keys (for instance description, keyword, location, etc.) have multiple values (e.g. "image showing a Gardenia" or "blooming flower") as shown in table 1. In general inverted

Table 1: Example for a multi field document describing a picture of a Gardenia

Field	Value
Description	"digital photo showing a Gardenia"
Tag	"Gardenia"
Tag	"Flower"
Tag	"Blooming"
Author	"Max Mustermann"
Rating	5
Date	July 2007

lists [Baeza-Yates and Ribeiro-Neto 1999] are used to allow retrieval optimized for speed. Image retrieval systems based on low level features store the features extracted from n indexed images as vectors $\mathbf{d_i}$ with $i \in \{1..n\}$ in a data base. In case of retrieval also a query is expressed as vector $\mathbf{q}$ and a metric is used to compute the distance between the query vector and all n vectors $\mathbf{d_i}$ in the data base. The result set R is then composed of the best matching images: $R = \{\mathbf{v}_i \| dist(\mathbf{q}, \mathbf{v}_i) \leq \epsilon\}$ whereas $\epsilon > 0$ denotes a treshold for a maximum distance and $dist(\mathbf{q}, \mathbf{v}_i)$ gives a measure for the relevance for ranking the results.

Main disadvantage of this approach is that it takes linear time (linear to the number of indexed images) to find the best matching images. In text retrieval performance regarding speed does not depend on the number of documents but on the number of terms, which is typically a smaller number than the number of documents in very large databases. A common approach in content based image retrieval is to use spatial access methods [Baeza-Yates and Ribeiro-Neto 1999] in addtion to a reduction of dimensions in the search space to speed up retrieval. Due to the difference between the approaches – inverted lists vs. spatial access methods – combined retrieval of high level metadata (especially textual) and low level (numerical) features is a challenging task.

Here, we confine our comparison onto the systems (probably) most relevant to our work in the field of metadata driven image retrieval (Flickr.com) and content based image retrieval (Caliph & Emir). Furthermore we include a short description of Magick, which is related to *Imagesemantics* in terms of feature combination. As IBM's Proventia Web Filter can not be actively queried for ad-hoc image retrieval the search results of the previous systems will be directly compared with our *Imagesemantics* system. For the sake of comparability, we perform our query in any system on the same kind of picture: A blooming flower

Figure 1: The kind of flower we are searching for – A blooming Gardenia

called *Gardenia* like the one shown in figure 1.

3.1 Flickr.com

Flickr is a typical Web 2.0 representative. It provides its users with functionalities to describe, tag and arrange images in web-based collections. Even more, as Flickr is a social software, the whole community might contribute to the stored images. Furthermore social networking in Flickr results in sub communities having different contexts and annotation behaviour. For that reason, the tags are frequently misleading as different users have different perspectives onto a certain picture or focus on different semantic levels (the image in general or a certain detail).

Similarly, images can only be retrieved via the images' metadata descriptions or tags. Thus, users can specify search terms. As tagging is a community wide process, the adequacy of search results is somewhat "arbitrary". In the case of our comparative search for a "Blooming Gardenia" our initial query "Gardenia" returned an unmanageable 3000 pictures, of which "only" 1.200 were explicitily tagged as "Gardenia". Therefore, we refined our query to "Blooming AND Gardenia" which returned a reasonable amount of 23 pictures only. However, the

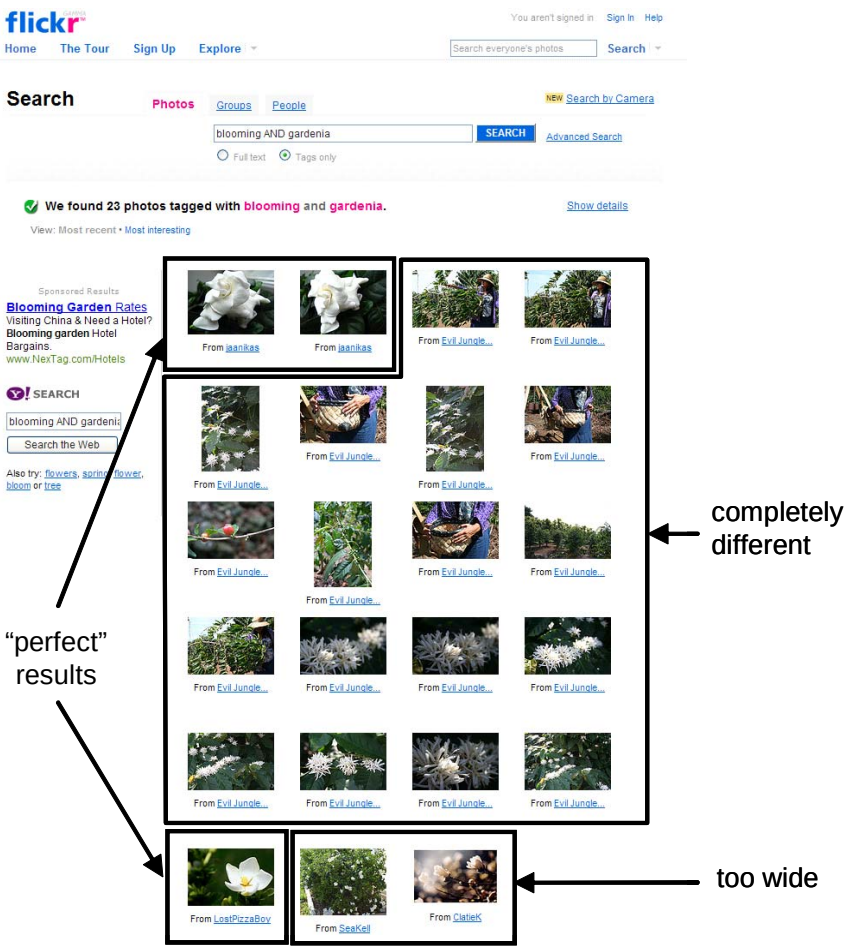

Figure 2: Flickr image search results for "blooming AND gardenia"

result set is quite disappointing (cf. figure 2). Our query returned only three pictures we were searching for (cf. figure 1), while the remaining 20 were quite different. Two of the images dealt with "Gardenia" but were "too wide" while the remaining 18 showed completely different pictures of coffee plants.

3.2 Caliph & Emir

Caliph & Emir are MPEG-7 based Java applications for image annotation and retrieval applying content based image retrieval techniques using MPEG-7 descriptors [Lux et al. 2003]. Besides extraction of existing information in digital photographs and transformation of these contents to MPEG-7, Caliph supports

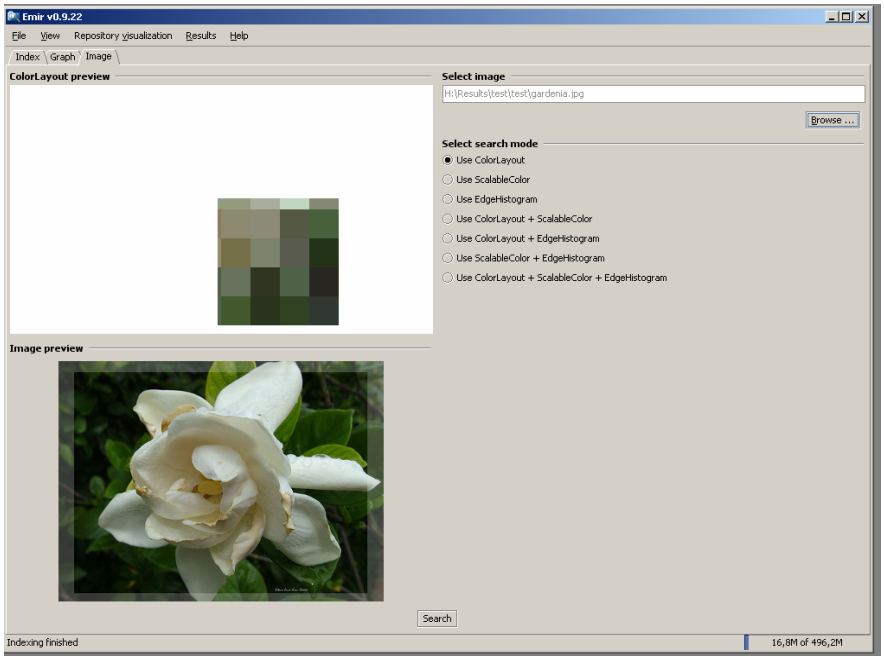

Figure 3: Initialization of a query in Caliph using the image of a *Gardenia*

the creation of semantic dependencies. For that purpose, the MPEG-7 descriptions in Caliph comprise metadata description, creation information, media information, textual annotation, semantics and visual descriptors [Lux et al. 2006].

On top of it, Emir supports content based image retrieval in local image repositories created with Caliph. The most sophisticated retrieval techniques applied in Emir are backed by MPEG-7 descriptors *Color Layout*, *Edge Histogram* and *Scalable Color*. Figure 3 shows the initialization of a query using the image of figure 1 as reference by applying the MPEG-7 *Color Layout* descriptor. For the sake of comparability, our reference query was evaluated on a local copy of all those 3.000 images of flickr, which contained "Gardenia‘" in their description. The 15 top-ranked images by Emir are shown on the right hand side of figure 4. Given the fact that a pre-selection of images based on their flickr-tags has manually been performed beforehand, the images retrieved are of much better quality than in the tag only search presented before. Nevertheless, about the half of the result size is quite different from what we have been querying for (lower right polygon on the right hand side of figure 4). Thus, there are two main drawbacks: First, Emir requires a manual preprocessing of Flickr images by Caliph in order to create a local collection. Second, the comparison of the

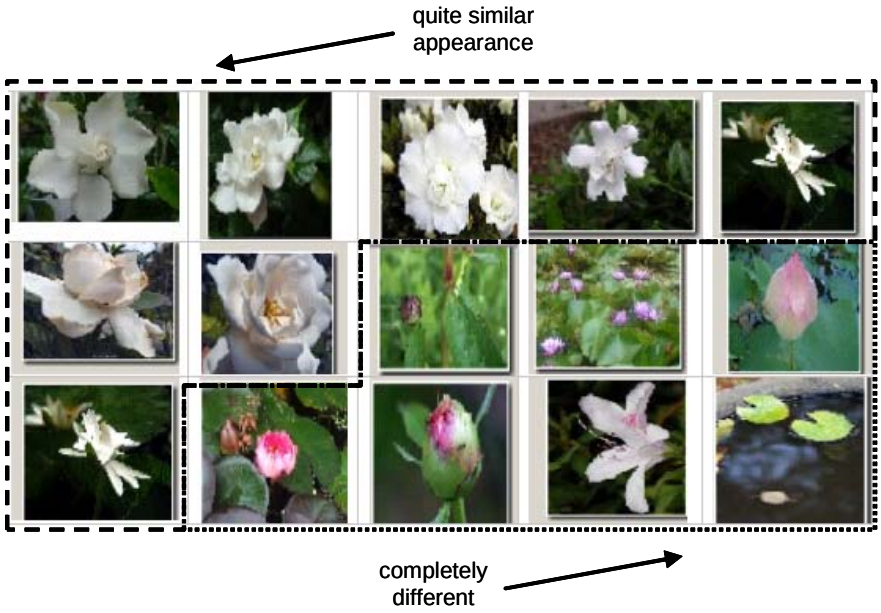

Figure 4: Query results retrieved from Caliph

reference image's feature vector with all the other image vectors takes very long.

3.3 Magick

The approach for of feature combination in *Magick* [Lux et al. 2004] is related to our work. Magick allows to cluster and visualize sets containing both textual and image data and displays the results in a 2D visualization (see fig. 5). Magick employs high level metadata (IPTC and EXIF in case of images and Dublin Core in case of text documents) as well as low level metadata (MPEG-7 low level features in case of images and term vectors in case of text documents) to compute pair wise distances between content items. The distance is normalized and a weighted combination of the different features is used to generate an overall distance matrix for hierarchical clustering and multidimensional scaling for the visual representation. Both clustering and MDS are highly sensitive to changes in the distance matrix, which itself heavily depends on the weights for the combination of different features. The actual selection of optimal weights is a complicated task and possibilities are manifold (as illustrated throught the weighting adjustment dialog in Magick in fig. 6). Therefore Magick is not able to support ad hoc retrieval tasks.

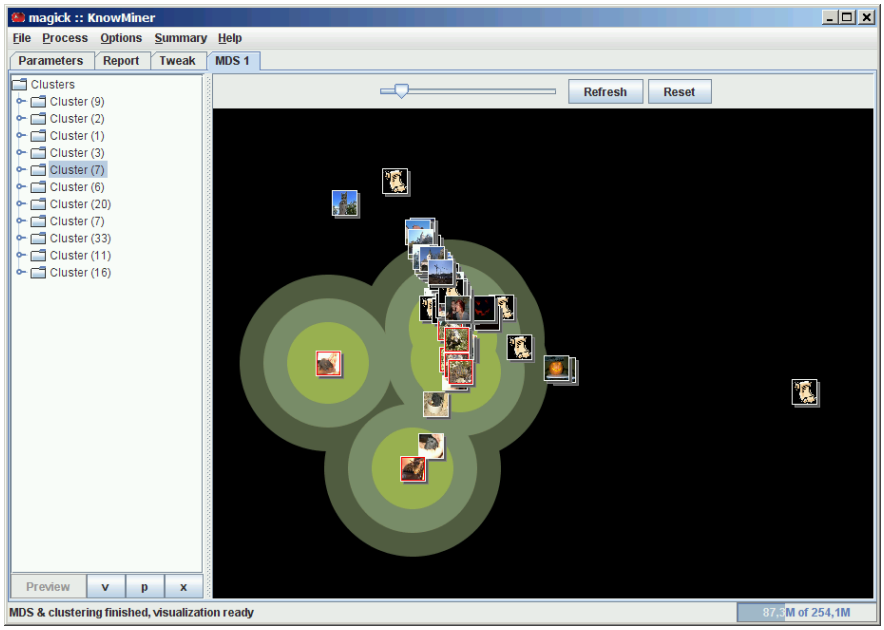

Figure 5: Visualization in Magick

3.4 Summary

The main challenge in image retrieval is a fast and concise interpretation of the query's semantic. As we demonstrated before, a solely text or content based analysis mostly leads to unsatisfying results. While the results obtained from a time consuming content based analysis performed on manually pre-selected images based on their Flickr tags proved to be much more precise, the ultimate goal seems to be the combination of an accelerated text and image analysis. Therefore, our *Imagesemantics* system links both approaches by a k-means clustering algorithm. By comparing the reference image with the cluster vectors, *Imagesemantics* allows its users a fast and concise opportunity to formulate search queries based on search terms and by specifying a reference image.

4 *Imagesemantics*: Rule-based Image Clustering & Retrieval

As we have presented in the previous chapter, solely text or content based image retrieval often leads to unsatisfying results. In order to close the so-called "semantic gap" our approach is a rule-based image clustering and retrieval system. Starting from our multimedia information systems MECCA [Klamma et al. 2005] and MEDINA [Spaniol and Klamma 2005] developed to foster the annotation, classification, discussion and thus, collaboration about multimedia contents

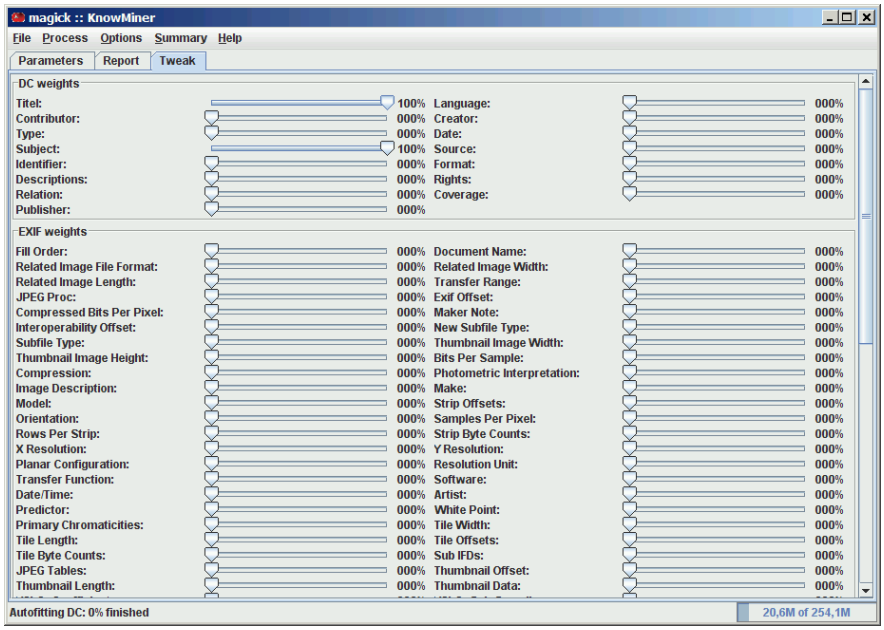

Figure 6: Weighting adjustment dialog in Magick

in communities of professionals, we explain how these high-level annotations can also be applied for the structured retrieval of contents.

4.1 Rule-based Image Clustering

Imagesemantics links text and content based image retrieval for a concise query processing. As already said, for speed-up reasons *Imagesemantics* is based on a k-means clustering algorithm. By comparing the reference image with the cluster vectors, this procedure image has not to be performed with all n images, but only k times with the reference vectors instead (keeping in mind that usually $n >> k$). Next, we step-by-step describe our rule-based clustering process (cf. figure 7).

In an initialization step *Imagesemantics* extracts the low-level feature vectors of a test collection of images. Here, we apply Flickr's Java API Jickr to obtain a relevant set of test images [Jickr 2007]. Subsequently, the feature vectors of the images are extracted. In order to ensure the interoperability of our data, we make use of the MPEG-7 metadata standard, particularly those having proven to perform concise and fast on image sets: *Edge Histogram*, *Scalable Color* and *Color Layout* (cf. section 2). From these feature vectors we create k image clusters, where k is arbitrary and can be defined upon initialization. The underlying

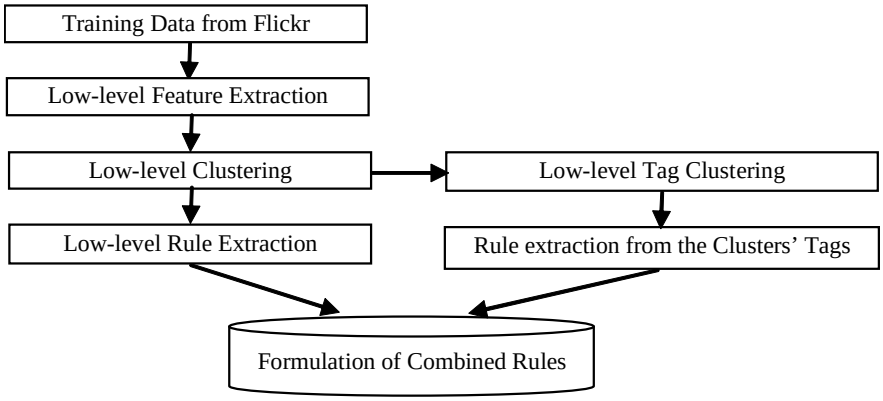

Figure 7: Rule-based image clustering process

algorithm is a modified k-means clustering process. In order to form preferably homogeneous clusters we apply Ward's minimum variance linkage [Fasulo 1999], in order to obtain the k cluster centroid feature vectors. In the next step, two operations are being performed. On the one side, low-level rules are extracted in order to express the maximum distance from a centroid's feature vector allowed for images belonging to it. On the other side, the members' tags are extracted as a tag-cloud of terms. The tag cloud vectors rules are derived for each cluster so that a sub-clustering based on the high-level semantic annotations takes place. In a final step, both low-level feature rules and high-level tag rules are combined. Thus, the gap between purely low-level content analysis and high-level metadata annotations can be bridged.

4.2 OWL-based Rules derived from Low-level Features and Tags

In the previous section we have explained how we extract the rules in our image clustering process. To describe the rules we are using some de facto standards in development of web-based information systems. The Resource Description Framework (RDF) [Beckett and McBride 2004] provides data model specifications and an XML-based serialization syntax. The Web Ontology Language OWL [Bechhofer et al. 2004] enables the definition of domain ontologies for many purposes like context modeling and sharing of domain vocabularies [Wang et al. 2004]. In the Semantic Web vision, OWL helps web services as agents to share information and interoperate [Chen et al. 2004]. Generally spoken, OWL has following usages: first, *domain formalization*: a domain can be formalized by defining classes and properties of those classes; second, *property definition*: individuals and assert properties about them can be defined; third, *reasoning*: one can reason about these classes and individuals. Using OWL an ontology is described

as a collection of RDF triples in the form of `(subject,predicate,object)`, in which `subject` and `object` are objects or individuals of an ontology and `predicate` is a property relation defined by the ontology. In order to make the previously extracted rules understandable and interpretable by a reasoner, we will now describe how the rules are represented in an OWL ontology stored in an eXist XML database [Meier, 2003].

Listing 1: Low-level feature rules

```
<owl:Class rdf:about="LowlevelCluster_k">
 <Centroid> Values </Centroid>
 <rdfs:subClassOf>
  <owl:Restriction>
   <owl:onProperty>
     <owl:ObjectProperty rdf:about="Distance"/>
   </owl:onProperty>
   <owl:allValuesFrom>
    <owl:Class rdf:about="Interval_kmin_kmax"/>
   </owl:allValuesFrom>
  </owl:Restriction>
 </rdfs:subClassOf>
</owl:Class>

<owl:DatatypeProperty rdf:about="Centroid">
 <rdfs:domain rdf:resource="LowlevelCluster_k"/>
 <rdfs:range
     rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

Listing 2: High-level tag rules

```
<owl:Class rdf:about="LowlevelCluster_k__Tagx">
 <Hastag>Tagname</Hastag>
 <rdfs:subClassOf rdf:resource="LowlevelCluster_k"/>
</owl:Class>

<owl:DatatypeProperty rdf:about="Hastag">
 <rdfs:range
     rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
 <rdfs:domain rdf:resource="_LowlevelCluster_k__Tagx"/>
</owl:DatatypeProperty>

<LowlevelCluster_k rdf:about="260407965_5c177d3703.mp7.xml">
 <rdf:type rdf:resource="LowlevelCluster_k__Tagx"/>
 <rdf:type rdf:resource="LowlevelCluster_k__Tagy"/>
</LowlevelCluster_k>
```

The class *LowlevelCluster_k* is the representative of the k^{th} cluster. This class contains the information about the centroid's feature vector as well as the cluster's range $Interval\_k_{min}\_k_{max}$ (cf. left hand side of figure 7). Based on these information it can now be specified whether an image belongs to a certain cluster

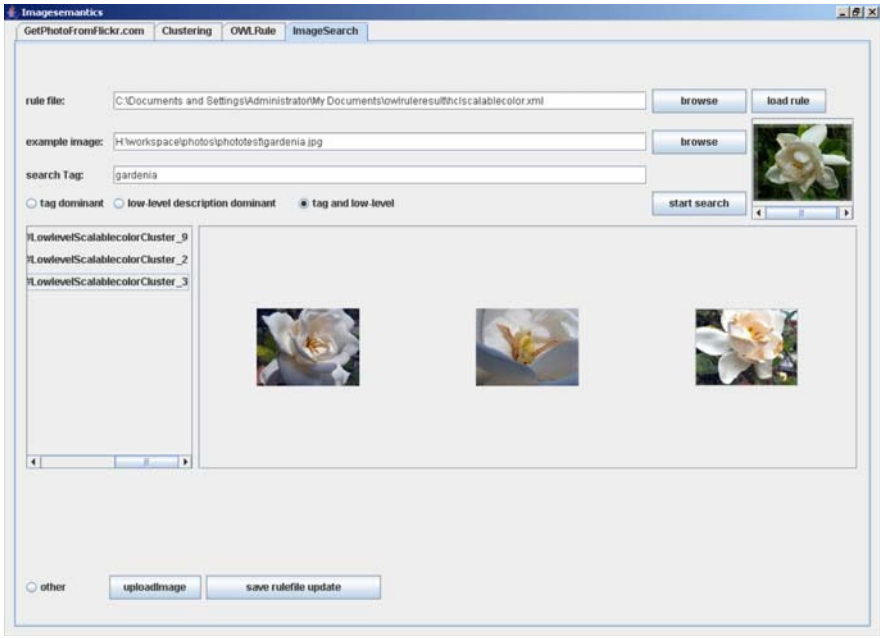

Figure 8: *Imagesemantics* search result based on low-level features and tags

or not. Similarly, the extracted rules from the clusters' tags can be expressed in OWL-classes. For instance, the class $LowlevelCluster\_k\_Tag_x$ contains the Tagname as a value of the x^{th} tag in cluster k (keeping in mind that each image and, thus, every cluster may be assigned with more than a single tag). As a result, for each cluster the associated high-level tag are formulated as a rule (cf. listing 1). In order to apply the inference mechanisms of an OWL reasoner, for each image an instance is being created (cf. listing 2).

In retrieval, the instances are in a first step being queried for a certain Tagname x. All those clusters are being identified, which contain this value. In our previous example the cluster $LowlevelCluster\_k$ would be one of the candidates. Then, the reference image's feature vector is being compared with the cluster's centroid vector. In case the difference is below a pre-defined threshold the dedicated cluster is prepared for the result. In a final step a selection takes place so that only those images of the chosen clusters will be shown, which are tagged by Tagname x (cf. figure 8).

5 Conclusions & Outlook

In this paper we have presented *Imagesemantics*, a system, which combines low-level and high level features for retrieval. Compared to a simple combination of low level and high level descriptors through weighted averaging of similarity functions, *Imagesemantics* relies on a cluster based index, which combines descriptors of both sides of the semantic gap. *Imagesemantics* is currently work in progress and a large scale evaluation is still missing, but first heuristic evaluations have shown that the results of our system are subjectively better than approaches solely relying on single level descriptors. Therefore, our system is a promising approach to narrow (or even bridge) the "semantic gap" between low-level content based retrieval and high-level metadata annotations in image retrieval. By comparing the search results of Flickr and Caliph & Emir, *Imagesemantics* shows that a combined approach can lead to more satisfying results.

In future, we will provide the functionalities of *Imagesemantics* via Web Services so that the system needs not to be used as a stand-alone Java application. In addition, we intend to enhance our Imagesemantics from solely image retrieval support to any other multimedia contents, particularly videos. In this aspect we focus our research on efficient processing mechanism for content-based key frame analysis and high-level textual descriptions. Finally, we plan to evaluate the accuracy and performance obtained by *Imagesemantics* on a larger scale. Hence, we want to compare our results with other systems based on standardized multimedia test sets we are currently developing in a community of professionals (`www.multimedia-metadata.info`) in the field of multimedia.

Acknowledgements

This work was supported by the German National Science Foundation (DFG) within the collaborative research center SFB/FK 427 "Media and Cultural Communication" and within the research cluster established under the excellence initiative of the German government "Ultra High-Speed Mobile Information and Communication (UMIC)". We thank Ming Cheng and our colleagues of the Multimedia Metadata Community (`www.multimedia-metadata.info`).

References

[Baeza-Yates and Ribeiro-Neto 1999] Baeza-Yates, R. A., Ribeiro-Neto, B. (1999). Modern Information Retrieval. Addison-Wesley Longman Publishing Co., Inc., 1999.

[Beckett and McBride 2004] Beckett D. and McBride,D. (2004) RDF/XML Syntax Specification (Revised), `http://www.w3.org/TR/rdf-syntax-grammar/` [10.7.2007].

[Bechhofer et al. 2004] Bechhofer, S. van Harmelen, F., Hendler, J., Horrocks, I. McGuinness, D. L., Patel-Schneider, P. F., Stein, L. A., Olin, F. W. (2004) OWL Web Ontology Language Reference, `http://www.w3.org/TR/owl-ref/` [10.7.2007].

[Chang et al. 2001] Chang, S.F., Sikora, T., Puri, A. (2001). Overview of the MPEG-7 standard. Special Issue on MPEG-7 IEEE Transactions on Circuits and Systems for Video Technology, IEEE, pp. 688-695.

[Chen 2006] Chen, M. (2006). Regelextraktion und -darstellung von Multimedia Semantiken. Diploma Thesis, RWTH Aachen University.

[Chen et al. 2004] H. Chen H., Finin T., Joshi, A. (2004) An Ontology for Context-Aware Pervasive Computing Environments in: Special Issue on Ontologies for Distributed Systems, Knowledge Engineering Review, 18(3):197-207

[DelBimbo 1999] Del Bimbo, A. (1999). Visual Information Retrieval. Morgan Kaufmann.

[Eidenberger 2004] Eidenberger, H. (2004). Statistical analysis of MPEG-7 image descriptions. ACM Multimedia Systems journal, 10, (2), pp. 84-97.

[Fasulo 1999] Fasulo, D. (1999). An Analysis of Recent Work on Clustering Algorithms. Technical Report 01-03-02, Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195.

[Flickr 2007] flickrTM (2007). `http://www.flickr.com/` [10.7.2007].

[Google 2007] Google Images (2007). `http://images.google.com/` [10.7.2007].

[IBM 2007] IBM Proventia Web Filter: Overview (2007). `http://www.iss.net/products/Proventia_Web_Filter/product_main_page.html`, [10.7.2007].

[ISO 2002] ISO ISO/IEC (2002). Information Technology - Multimedia Content Description Interface - Part 3: Visual. ISO/IEC 15938-3:2002, ISO.

[ISO 2003] ISO/IEC (2003). Information technology - Multimedia content description interface - Part 5: Multimedia description schemes. ISO/IEC 15938-5:2003, ISO.

[Jickr 2007] Jickr - Flickr Java API (2007). `https://jickr.dev.java.net/` [10.7.2007].

[Klamma et al. 2005] Klamma, R., Spaniol, M., Jarke, M. (2005). MECCA: Hypermedia Capturing of Collaborative Scientific Discourses about Movies. informing science: The International Journal of an Emerging Discipline, 8, pp. 3 - 38.

[Lux et al. 2003] Lux, M., Becker, J., Krottmaier, H. (2003). Semantic Annotation and Retrieval of Digital Photos. In Proc. of CAiSE '03 Forum, `http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-74/files/FORUM_22.pdf` [10.7.2007].

[Lux et al. 2004] Lux, M., Granitzer, M., Kienreich, W., Sabol, V., Klieber, W. , Sarka, W. (2004). Cross Media Retrieval in Knowledge Discovery. In: Proc. of PAKM 2004, Vienna, Austria, December `http://mathias.lux.googlepages.com/mlux-pakm04-preprint.pdf` [18.5.2007].

[Lux et al. 2006] Lux, M., Klieber, W., Granitzer, M. (2006). On the Complexity of Annotation with the High Level Metadata. J.UKM, 1, (1), pp. 54-58.

[Meier, 2003] Meier, W.(2002). eXist : An Open Source Native XML Database. In: Web, Web-Services, and Database Systems, NODe 2002 Web and Database-Related Workshops, Erfurt, Germany, October 7-10, Revised Papers, volume 2593 of LNCS, Springer-Verlag, Berlin Heidelberg, pp. 169 - 183.

[Robertson et al. 2004] Robertson, S., Zaragoza, H., Taylor, M. (2004). Simple BM25 extension to multiple weighted fields. In: Proc. of CIKM '04, ACM Press, New York, NY, USA, 2004, pp. 42-49.

[Smeulders et al. 2000] Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R. (2000) Content-based image retrieval at the end of the early years. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2000, 22, 1349-1380

[Spaniol and Klamma 2005] Spaniol, M., Klamma, R. (2005). MEDINA: A Semi-Automatic Dublin Core to MPEG-7 Converter for Collaboration and Knowledge Management. In: Multimedia Repositories. In Proc. of I-KNOW '05, Graz Austria, J.UCS Proceedings, LNCS 1590, Springer-Verlag, pp. 136 - 144.

[Wang et al. 2004] Wang, X. H. Zhang, D. Q., Gu, T., Pung, H. K. Ontology Based Context Modeling and Reasoning using OWL in: PERCOMW '04: Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops, Washington, DC.: IEEE Computer Society, p. 18

[Yahoo 2007] Yahoo! Photos (2007). `http://photos.yahoo.com/` [10.7.2007].