A Hybrid Ontology and Content-Based Search Engine For Multimedia Retrieval*

Charalampos Doulaverakis, Evangelia Nidelkou, Anastasios Gounaris, and
Yiannis Kompatsiaris

Informatics and Telematics Institute
Centre for Research and Technology Hellas, Thessaloniki
e-mail: {doulaver,nidelkou,gounaris,ikom}@iti.gr

Abstract. Huge amounts of digital visual content are currently available, thus placing a demand for advanced multimedia search engines. The contribution of this paper is the presentation of a search engine that is capable of retrieving images based on their keyword annotation with the help of an ontology, or based on the image content to find similar images, or on both these strategies. To this end, the search engine is composed of two different subsystems, a low-level image feature analysis and retrieval system and a high-level ontology-based metadata structure. The novel feature is that the two subsystems can co-operate during the evaluation of a single query in a hybrid fashion. The system has been evaluated and experimental results on real cultural heritage collections are presented.

1 Introduction

Multimedia content management plays a key role in modern information systems. From personal photo collections to media archives, cultural heritage collections and bio-medical applications, an extremely valuable information asset is in the form of images and video. To provide the same functionalities for the manipulation of such visual content as those provided for text processing, the development of search engines that perform the retrieval of the material is of high significance. Such an advanced, semantic-enabled image search engine is the subject of the work presented in this paper.

To date, two main approaches to image search engine techniques have been proposed, annotation-based and content-based. The former is based on image metadata or keywords that annotate the visual content. A well known example that falls into this category is Images Google Search[1]. The metadata that a search engine of this kind typically relies on refers either to the properties

[1] http://images.google.com/

of the image itself or to its content. Examples of image properties include the name of the image file, its creation date, copyright information, image format, resolution and so on. On the other hand, content metadata correspond to the properties of the entities depicted, such as persons and objects. Several variants of annotation-based multimedia search engines have been proposed that assume manual annotation (e.g., [6]) or they provide support for automatic annotation (e.g., [14]). This search approach has benefitted significantly from the advances in the Semantic Web and the ontologies (e.g., [13]). Ontologies are *"an explicit specification of a conceptualisation"* [7], and they guarantee firstly a shared understanding of a particular domain, and secondly, a formal model that is amenable to unsupervised, machine processing. The use of ontologies has also made possible the integration of different content under a unified description base where various collections can be accessed using a common querying framework. For example, some museums use ontologies for storing and describing their collections, so that users can browse and explore the museum collections, and understand the way in which the items are described and organized. Indicative examples of such systems are Artefacts Canada[2] and MuseoSuomi[3].

However, metadata-based search is often insufficient when dealing with visual content. To tackle this, a second complementary approach has been employed: content-based search. The core idea is to apply image processing algorithms to the image content and extract low-level features; the retrieval is performed based on similarity metrics attempting to imitate the way humans perceive image similarity (e.g., [12]). This approach allows the user to retrieve images that are similar in terms of the entities depicted but fails to capture the underlying conceptual associations. A well-known example is the ImageSeeker tool[4].

This paper focuses on a novel search engine that is capable of performing not just these two approaches to image search but also to combine them in a novel, hybrid way. Moreover, the search engine has been employed in (and motivated by) a real scenario, which involves the development of advanced techniques to access multimedia cultural heritage material. Cultural heritage collections are usually accompanied by a rich set of metadata, and thus are suitable for our case. The paper also provides insights into the performance and the efficiency of each strategy.

The remainder of the paper is structured as follows. Section 2 describes the broader cultural heritage management project that has motivated the development of the search engine presented. The content-based, ontology-based and hybrid flavors of that search engine are the topic of Section 3. Insights into the performance of the different approaches appear in Section 4. Section 5 deals with the related work and Section 6 concludes the paper.

---

[2] http://www.chin.gc.ca/English/Artefacts_Canada/

[3] http://www.museosuomi.fi/

[4] http://www.ltutech.com

2   A Cultural Heritage Use Case

REACH is an ongoing project, the objective of which is to develop an ontology-based representation in order to provide enhanced, unified access to heterogeneous cultural heritage digital databases. The system under development integrates Greek databases of cultural content offering unified access and efficient methods of search and navigation of content to users, while enabling commercial exploitation. The complete system is composed of the following subsystems: (i) a cultural heritage web portal for unified access to the information and services, (ii) a digitalization system for the efficient digitalization of artwork and collections, (iii) an ontology to describe and organize cultural heritage content, (iv) a multimedia content-based, as well as ontological-based, search engine to offer advanced choices of searching methods, (v) an e-Commerce section for the commercial exploitation of the portal. The main content provider for the project is the Centre for Greek and Roman Antiquity (KERA) who offers a large collection of inscriptions and coins from the Greco-Roman time period, accompanied with detailed documentation. This paper focuses on the fourth subsystem.

3   Multimedia Retrieval

Within the search engine being developed in the framework of REACH, the user has the ability to initiate a retrieval procedure by using one of three different available options. These options deploy state of the art and novel technologies of Information Retrieval for performing multimedia retrieval in a way that there is a better chance for a user to find desired content. These options namely include: a) content-based retrieval, b) ontology-based retrieval, and c) hybrid retrieval which makes a combined use of the two aforementioned methods.

3.1   Content-based Multimedia Retrieval

By utilizing this option, users are able to perform a visual-based search by taking advantage of low-level multimedia content features. Content-based search is more appropriate for the cases where users feel that they can provide prototype multimedia content which is similar to the content they are looking for. The search engine can handle 2D still image and video, and 3D models. A user is able to provide, as the input query, an example of the multimedia content she is interested in, and, based on the extracted descriptors of the input and the stored offline-generated descriptors of the content repository, the system performs a visual similarity-based search and relevant results are retrieved.

For proper handling of the various content types, different strategies are employed for each type in the offline analysis process. This is explained in more detail in the following subsections.

2D Still Image Analysis Analysis of 2D images is performed in a two-step fashion. To enable meaningful region detection in the available cultural heritage images collections, a segmentation process takes place using the approach described in [11]. There are several advantages in using regions for image retrieval and these are mainly derived from the fact that users usually search for objects displayed in images rather than whole images instead. This is the typical case in the area of cultural heritage as the main interest in retrieval is the item being displayed in an image regardless of any surroundings or background.

The second step in analysis involves low-level feature extraction from the resulting regions of the segmentation mask and also from the whole image itself. For this purpose, the MPEG-7 features were selected as they represent the state of the art in low-level visual descriptors. For the extraction, the MPEG-7 eXperimentation Model (MPEG-7 XM) [1] was used as it realizes the standardized descriptors and apart from extraction it also utilizes methods for similarity based retrieval. The extracted descriptors are binary encoded into bitstreams and stored in a separate database. Eight descriptors are used in total, namely *Color Layout*, *Color Structure*, *Dominant Color*, *Scalable Color*, *Edge Histogram*, *Homogeneous Texture*, *Contour Shape* and *Region Shape*. These descriptors complement each other and are adequate for describing the object appearance in detail.

During the retrieval process, since more than one descriptors are used and due to the inter-variation of their dimensionality and range, the overall distance to the description of the query image is calculated using normalized distances for each descriptor. In the last step, the retrieved images are gathered and displayed as results.

Video Analysis For video analysis, the video stream is firstly divided into shots using the method described in [9]. For each detected shot, a keyframe is extracted which is treated as a compact representation of the entire shot. This keyframe is then analyzed as in the still image case, i.e., it is segmented into regions and feature extraction using MPEG-7 XM is performed.

During retrieval, MPEG-7 XM is employed as in the previous case, and the relevant shots are returned as result, based on their keyframe similarity. By adopting this strategy for video analysis, the video retrieval process is reduced to an image retrieval scenario. Such a strategy has been followed successfully in the TRECVID Video Retrieval Evaluation and its results are very promising [10].

Fig. 1 summarizes the analysis for both Image and Video.

3D Content Analysis For 3D content analysis, an approach similar to the one described in [4] is followed. The VRML representation of the 3D model is firstly manipulated in such a way so that its mass center is placed on the zero point of a 3D orthogonal coordinate system. Furthermore, the model is scaled so that the maximum distance of a voxel to the center of mass equals to one. Subsequently, a generalized 3D Radon transform is used for extracting feature vectors. To
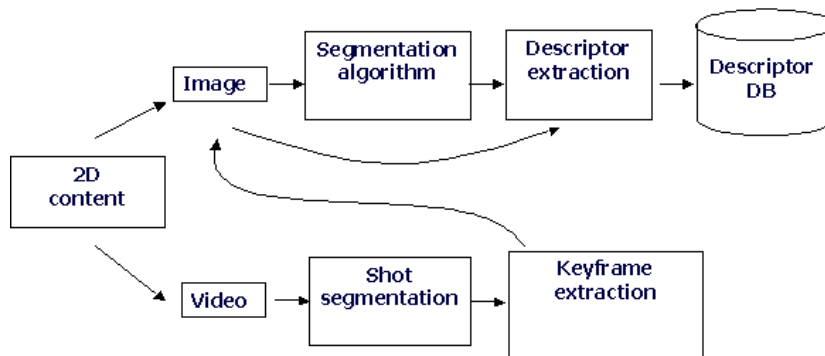
Fig. 1. 2D multimedia content analysis process.

decrease the vector size, dimensionality reduction techniques are employed. The retrieval is performed, as in the previous cases, on the grounds of the similarity distance; to this end the euclidian distance of the vectors is calculated.

3.2 Ontology-based Multimedia Retrieval

Cultural heritage collections are accompanied by a rich set of metadata that describe various details related to each item regarding historical data or details regarding administrative information like, for example, current exhibition location. However, this metadata is often unstructured or registered in a non-standard form, usually proprietary, for every collection, which renders them unusable for inter-collection searching. To overcome this problem, appropriate ontologies for the cultural heritage domain have been defined inside the scope of REACH. Such ontologies (both their definition and their instantiation) can be used for searching purposes when the search criteria are the collection item metadata rather than the visual appearance, as in the previous case.

Taking into account the content originally available, namely the Inscriptions and Coins collections of KERA, an ontology infrastructure has been defined to efficiently describe and represent all knowledge related to each collection. The proposed architecture makes use of three different ontologies, namely *Annotation*, *Coins* and *Inscriptions*. These ontologies are detailed below.

A large set of information fields inside the metadata is common for each item, regardless of the collection that is part of. As such, it was decided to use a separate ontology specifically intended for holding this data, which include information like date and place of creation, current location, construction material, dimensions, etc. Such data is an example of properties that appear in and characterize every item inside the collection. Consequently, the role of the *Annotation* ontology is to conceptualize and hold all common data in a structured

way, thus forming a representation standard for every collection to be integrated with the search engine.

The properties that are specific to a collection item category are captured by other ontologies; more specifically there is a separate ontology for each category, as the particular details that correspond to the collection items can vary greatly for each class. An example is the Coins and Inscriptions metadata. The information that one requires to search through Coin collections, such as monetary subdivision, is significantly different from the information used for Inscriptions searching (e.g, inscription text). A thorough study of the metatada has shown that this kind of specific information does not overlap across items as is the case with the *Annotation* ontology. As a result, the definition of a *Coins* and an *Inscriptions* ontology was the most appropriate approach in our case, since it can efficiently handle the data. Moreover, it does not restrict the extensibility of the system as the addition of cultural items of an additional type will only require the definition of a specific domain ontology for that type, and the mapping of its common data to the *Annotation* ontology.

As a further step to support interoperability of the system with other semantic-enabled cultural heritage systems, the aforementioned ontologies were mapped to the CIDOC-CRM [5] core ontology which has been proposed as an ISO standard for cultural heritage material structuring and representation. To enable this functionality, the CRM was thoroughly studied and appropriate mappings between the concepts of our defined ontologies and the CRM were appointed.

During search time, the system uses all three ontologies so that semantically connected items can be retrieved according to users selections. For example, the system can automatically retrieve items that were made of the same material, or in the same period or were found in the same place, and so on.

Illustrations of the developed search engine are displayed in Fig. 2. Using the infrastructure described, a user can initiate a search process looking for items with specific characteristics that are captured by the ontologies. As it can be seen from Fig. 2a the GUI provides a view of the three ontologies; selected concepts of each one are automatically organized according to their class hierarchy in a tree-like fashion. In the example in the figure, a search for the available bibliographic references has been requested by selecting the appropriate class from the ontology and, as filtering predicates, the user has selected items that are exhibited in a specific museum and are referenced in a specific historical book. The results are displayed as shown in Fig. 2b.

3.3 A Hybrid search engine

Often the user is interested in items that are both visually and semantically similar. With a view to supporting such functionality, the hybrid search engine provides a novel retrieval method, in which both visual and ontology search is employed for the same query. This novel method automatically combines different types of search results, and complements content-based search with ontology-based search and vice versa. It is important to note that the hybrid engine generates the new queries involved to retrieve more results in a way transparent

(a)   (b)

Fig. 2. a) Search interface where the user can select one or more ontology instances and concepts to form the search criteria, b) the results of the ontology-based search for specific selections.

to the user. The final result sets are integrated and are presented to the user in a unified manner.

The whole process is illustrated in Fig. 3. Let us assume that the user initiates a query by providing an example, as in the case of simple content-based search. This is depicted as case A in Fig. 3 and comprises of three steps. In the first step ($1A$) the content-based search is completed by analysing the provided multimedia content (i.e., performing the segmentation, extracting the low-level MPEG-7 descriptors and evaluating the distance between the prototype and the other figures stored in the multimedia database). The second step ($2A$) takes into account the metadata (which are mapped to the relevant ontologies) of the highest ranked results. For instance, the system may detect that the most common ontology concept of the highest ranked results in terms of visual similarity is the creation date or the place of exhibition of such results. Based on this information, an ontology-based query is formulated internally in the search engine, which contacts the knowledge base and enriches the result set with multimedia content that is close semantically to the initial content-based results ($3A$). Consequently, the response returned to the user covers a wider range of items of interest, thus facilitating the browsing through the collection and placing the burden of composing queries to the system instead of the user.

The reverse process is equally interesting (case B in Fig. 3). Here, the initial query is a combination of terms defined in the ontology, e.g., *"Artefacts from the 1st century BC"*. The knowledge base storing the ontology returns, as the first step ($1B$), the items that fall into that category. The second step ($2B$) involves the extraction and clustering of the low-level multimedia features of this initial set. In other words, the system detects the dominant color, the shape, the texture,
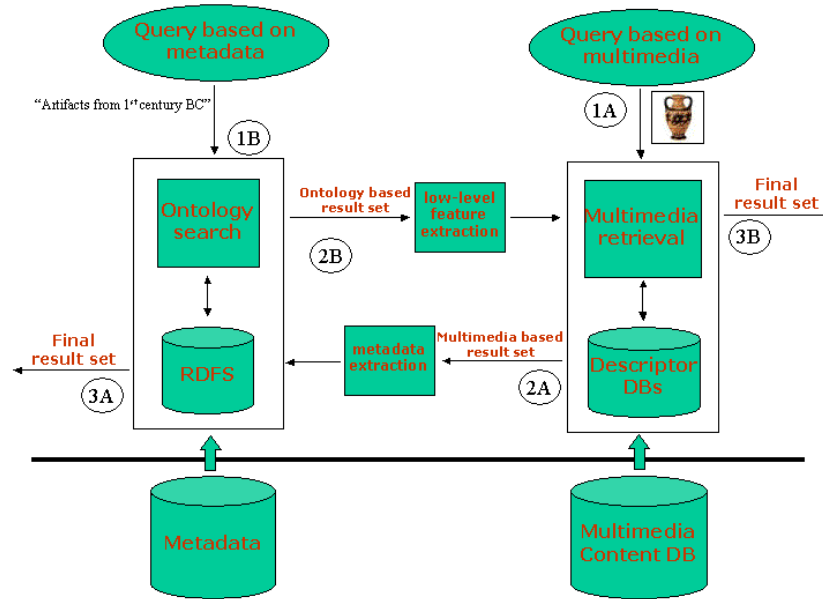
Fig. 3. Hybrid search engine architecture.

etc., of the larger set of the results of the the first step. As discussed previously, these features essentially drive the content-based search, which is performed in the final step (*3B*). Again, this leads to a more complete result set. Note that the hybrid search initiated by a query on the ontology is under development.

Fig. 4 provides an example of a hybrid search, in which the initial content-based search is coupled with an ontology-based one, thus providing a more complete result set. Fig. 4a shows the results for a specific content-based search, i.e., when only visual similarity has been taken into account. Fig. 4b illustrates how the initial results are enriched when an ontology-based search is triggered transparently to the user. In this example, the concept used for the ontology-based search is the"Date of Creation", which was found to be most common among the first initial content-based results.

4   Evaluation

In the previous section, three different search policies were presented that provide three complementary options to query formulation, so that the users can find their desired content even in the case where the search criteria are rather complex. In this section, we present a closer inspection on the efficiency and the performance of each of the two basic methods (content- and ontology-based)
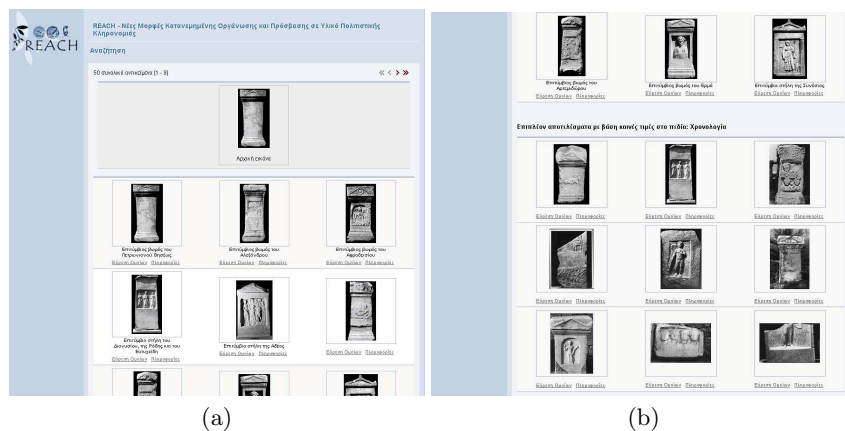
Fig. 4. a) Content-based results for the image at the top, b) Hybrid-based results for the same image query. The hybrid results were based on the frequency of common values for the "Date" class.

and draw conclusions on the advantages and disadvantages of each method in respect to precision of retrieval and response times.

The experiments were conducted on a PC, with a P5 3.0GHz Intel CPU and 1GB RAM. The knowledge base containing the ontological metadata is Sesame 1.2 running a MySQL DBMS at the back-end. MySQL is also used to store the actual non-multimedia content and the links to the multimedia files. The dataset is consisted of roughly 200 inscription images along with a complete set of metadata information. The descriptors are stored in a collocated MPEG-7 XM server. For content-based and ontology-based search, five queries, either visual or semantic, were used and the mean times are presented below. To evaluate the content-based search we selected five random images of inscriptions. The five semantic query tasks were: (i) "Find the items that are dated in the 3rd century BC"; (ii) "Find all inscriptions"; (iii) "Find the items that are referred to in a specific book and are exhibited in the Museum of the City of Veroia"; (iv) "Find the inscriptions with a specific ancient greek text"; and (v) "Find the items that were found in a specific location and are dated in the 1st or 2nd century AD".

Fig. 5 shows the Precision-Recall diagram for each one of the two methods, content-based and ontology-based retrieval. The curves correspond to the mean precision value that was measured after the five retrieval tasks. The precision of the ontology-based method is, as expected, 100% as we assume that the complete set of the metadata related to the items is manually edited and is precise, whereas the precision of the content-based depends on the efficiency of the distance evaluation algorithm. As such, this method is not as satisfactory, in terms of precision, as the ontology-based search. This is mainly due to the available
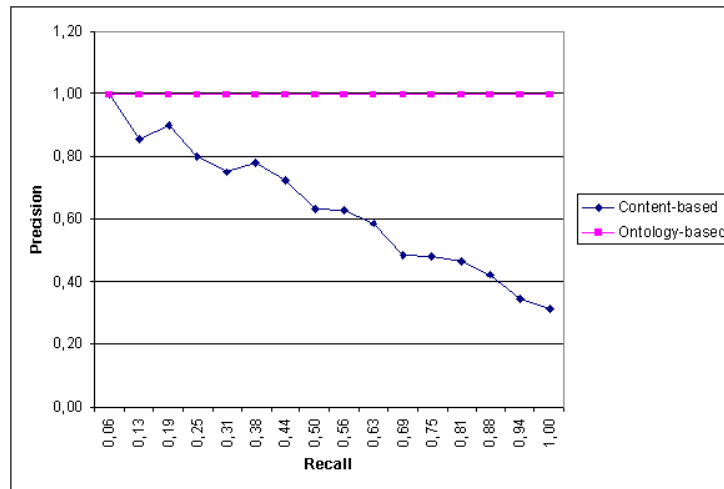
Fig. 5. Precision-Recall diagram for the ontology- and content-based methods

visual content, which is characterized by a rather small variance in terms of structure, i.e., all inscriptions have roughly the same shape apart from those who have sustained damage and have somehow lost their original shape. The ground truth used in the content-based experiments was the (subjective) visual similarity of the "inscriptions", i.e., basically their shape which can be either rectangular, or oval or undefined for the broken ones.

The average response times for each of the two retrieval methods are illustrated in Table 1, where it is evident that the ontology-based retrieval is much faster than the full content-based one. However, we should note that the ontology-based search depends on the availability of metadata in the form of ontology instantiations. If we assume that the metadata used in the content-based search (i.e., the visual features) are evaluated and stored in the preprocessing step in the multimedia database instead of the MPEG-7 XM server, then the time cost of content-based search is reduced to 0.068 sec, outperforming the ontology-based approach. This time is needed for strict image matching and feature extraction; however, because the MPEG-7 XM is running as a server, a relatively large amount (0.7sec approx.) of time is spent in socket communication and parsing, which adds a bottleneck to performance. Another characteristic of the content-based search is its scalable behavior: when the dataset grows larger, response times increase in a sub-linear manner. Experimental results with the MPEG-7 XM in server mode and a dataset of 2000 images have shown an average response time of 1.02 sec, i.e., a tenfold increase in the dataset corresponds only to a 32% increase in response time.

Method	Time
Content-based including the communication cost	0.773 sec
Content-based without the communication cost	0.068 sec
Ontology-based	0.163 sec

Table 1. Response times for ontology and content-based retrieval

The behavior of the hybrid search is expected to combine the benefits of the other two approaches for some queries. However, we do not present Precision-Recall graphs as these strongly depend on the nature of the retrieval task, and more specifically, on whether such a task can benefit from the combination of visual features and ontology concepts (e.g., "find all coins that are either similar to the one provided or that were created in the 1st century BC"). As such, solid measurement method for the hybrid search is difficult to obtain because of the strongly subjective nature of this proposed search option.

Commenting on the results of the two methods, it is evident that ontology-based search has better performance in both precision and time when compared to the full content-based search. Nevertheless, we should keep in mind that (i) these methods work on different representations of the available data and their use is intended to satisfy different needs; and (ii) ontology-based search presupposes the manual annotation of collection items. In summary, ontology-based search aims at making use of the metadata associated to an item, with respect to historical data (e.g., date of creation, place, etc.), while the content-based search aims at making use of lower level characteristics of the multimedia content corresponding to an item, like shape and color distribution, which can be automatically extracted. Such information is not likely to be found in the meta-data escorting a cultural heritage collection. Someone looking for items that are similar in shape, for instance, will use visual similarity as compared to a user interested in finding items that belong to a certain time period, and thus benefits more from the semantic search engine. Hybrid search is proposed as a heuristic way of combining the above two methods to provide results sets that could potentially be of relevance, and are based both on visual features and on the concepts defined in the ontology.

5 Related Work

Multimedia search engines have attracted a lot of interest both from the web search engine industry (such as Google, Yahoo!, and so on) and from academia. Also, the emergence of MPEG-7 standard has played a significant role in content-based search becoming a mature technology. For a survey, the reader can refer to [2, 3]. However, to the best of the authors' knowledge, no framework has been proposed on the combined use of ontology- and content-based retrieval.

In the domain of cultural heritage information retrieval and management using semantic technologies, there are two notable efforts. MuseuoSuomi [8] uses

facets as an intuitive user interface to effectively drive the user to follow a well defined path within the ontologies for retrieving specific cultural items. The path is formed by cross-querying all the underlying ontologies so that combinational queries can be made. Although the system handles metatata efficiently, there is no support for content-based multimedia retrieval. On the other hand, the SCULPTEUR project [13] makes use of the CIDOC-CRM to enable concept-based browsing of the metadata. In addition, SCULPTEUR employs content-based search for both 2D images and 3D models using proprietary methods and descriptors. However, the two methods cannot be combined. Finally, the search facility of the Hermitage in St. Petersburg[5] employs both ontology-based and content-based techniques, but these techniques are not are amalgamated as in our case.

6 Conclusions

In this paper, a novel search engine was presented for effectively searching through multimedia content (2D/3D image and video) related to cultural heritage collections. The engine adopts three methods for retrieval: two autonomous and one combinational. The ontology-based method makes use of the semantic mark-up metadata accompanying each collection where an illustrative user interface is used for graphical query formulation. The content-based method makes use of the low-level visual characteristics of the multimedia material while the hybrid method, which is the main contribution of this work, makes a combined use of the previous two methods for offering a more complete result set to the user.

Future work includes the extension of the hybrid search engine and the integration of additional cultural content. Finally we are investigating the addition of a semantic recommendation engine to be able to make additional query suggestions to the user in an automatic manner.

References

1. MPEG-7 eXperimentation Model
   . *http://www.lis.ei.tum.de/research/bv/topics/mmdb/e_mpeg7.html.*
2. Gwendal Auffret, Jonathan Foote, Chung-Shen Li, Behzad Shahraray, Tanveer Syeda-Mahmood, and HongJiang Zhang. Multimedia access and retrieval (panel session): the state of the art and future directions. In Shih-Fu Chang, editor, *MULTIMEDIA '99: Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, pages 443–445, New York, NY, USA, 1999. ACM Press.
3. Shih-Fu Chang, Qian Huang, Thomas Huang, Atul Puri, and Behzad Shahraray. Multimedia search and retrieval. In A. Puri and T. Chen, editors, *Advances in Multimedia: Systems, Standards, and Networks.* New York: Marcel Dekker, 1999.
4. P. Daras, D. Zarpalas, Tzovaras D., and M.G. Strintzis. Efficient 3-D Model Search and Retrieval Using Generalized 3-D Radon Transforms. *IEEE Transactions on Multimedia*, 8(1):101–114, 2006.

---

[5] http://www.hermitagemuseum.org

5. M. Doerr. The CIDOC-CRM An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3):75–92, Fall 2003.

6. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The QBIC system. *IEEE Computer*, 28(9):23–32, 1995.

7. T.R. Gruber. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5:199–220, 1993.

8. E. Hyvonen, M. Junnila, S. Kettula, E. Makela, S. Saarela, M. Salminen, A. Sreeni, A. Valo, and K. Viljanen. Finnish Museums on the Semantic Web. User's Perspective on MuseumFinland. In *Museums and the Web 2004, Virginia USA*, March 2004.

9. V. Kobla, D.S. Doermann, and K.I. Lin. Archiving, indexing, and retrieval of video in the compressed domain. In *Proc. SPIE Conference on Multimedia Storage and Archiving Systems, vol. 2916*, pages 78–89, 1996.

10. V. Mezaris, H. Doulaverakis, S. Herrmann, B. Lehane, N. O'Connor, I. Kompatsiaris, and M.G. Strintzis. Combining Textual and Visual Information Processing for Interactive Video Retrieval: SCHEMA's Participation in TRECVID 2004. In *TRECVid 2004 Workshop, Gaithersburg, Mairyland, USA*, November 2004.

11. V. Mezaris, I. Kompatsiaris, and M. G. Strintzis. Still Image Segmentation Tools for Object-based Multimedia Applications. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(4):701–725, June 2004.

12. O. Mich, R. Brunelli, and C. Modena. A survey on video indexing. *Journal of Visual Communications and Image Representation*, 10:78–112, 1999.

13. P. Sinclair, S. Goodall, P.H. Lewis, K. Martinez, and M.J. Addis. Concept browsing for multimedia retrieval in the SCULPTEUR project. In *Multimedia and the Semantic Web, held as part of the 2nd European Semantic Web Conference, Heraklion, Crete, Greece*, May 2005.

14. A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.