

Hybrid Ontology and Visual-based Retrieval for Cultural Heritage Multimedia Libraries

Stefanos Vrochidis, Charalambos Doulaverakis, Lambros Makris, Anastasios Gounaris,
Evangelia Nidelkou, Ioannis Kompatsiaris and Michael G. Strintzis
Informatics and Telematics Institute
Thessaloniki, Greece
{stefanos, doulaver, lmak, gounaris, nidelkou, ikom, strintzi}@iti.gr

Abstract

Nowadays, an increasingly growing demand for the creation of digital multimedia libraries is arising, as huge amounts of digital visual content are becoming available. The content that resides in these libraries should be easily retrievable and classified in order to be fully accessible. The contribution of this paper is the introduction of a hybrid multimedia retrieval model accompanied by the presentation of a search engine that is capable of retrieving visual content cultural heritage multimedia libraries as in three modes: (i) based on their semantic annotation with the help of an ontology; (ii) based on the visual features with a view to finding similar content; and (iii) based on the combination of these two strategies. To achieve this, the retrieval model is composed of two different parts, a low-level visual feature analysis and retrieval and a high-level ontology infrastructure. The main novelty is the way in which these two co-operate transparently during the evaluation of a single query in a hybrid fashion, making recommendations to the user and retrieving content that is both visually and semantically similar.

1. Introduction

Digital multimedia Libraries are organized in order to store and manage huge amounts of audiovisual content for universal remote access.

The huge amount of visual digitized information has to be structured and annotated in the proper way in order to be accessible and retrievable. To provide functionalities for the manipulation and knowledge retrieval from such visual content, a key aspect is the development of more efficient search engines to handle image and video files. To date, two main approaches to image search engine techniques have been proposed, namely annotation-based and

content-based. Several variants of annotation-based multimedia search engines have been proposed. Some of them assume manual annotation reducing visual search to information retrieval [1], while others provide support for automatic annotation.

This search approach has benefitted significantly from the advances in the Semantic Web and ontologies (e.g., [2]), so that annotations can have well-defined semantics. Ontologies are “an explicit specification of a conceptualization” [3], and they guarantee firstly a shared understanding of a particular domain, and secondly, a formal model that is amenable to unsupervised, machine processing.

However, annotation and semantic-based search are often insufficient when dealing with visual content. To tackle this problem, a second complementary approach has been devised: content-based search. The core idea is to apply image processing and feature extraction algorithms to the visual content and extract low-level visual features, such as color layout and edge histogram [4].

This paper focuses on a hybrid retrieval model by combining in an novel way the content and annotation based approaches. A search engine has been developed implementing this model using images from the culture domain. Regarding content-based search, the engine employs state-of-the-art techniques, which involve automatic segmentation of 2D visual content and MPEG-7 features extraction, while the novel, hybrid search functionality is capable of extending either the content-based search by making user suggestions of additional, potentially interesting results.

The remainder of the paper is structured as follows. Section 2 introduces the retrieval model while the evaluation procedure including the presentation of the search engine is presented in Section 3. Examples and results of the hybrid engine appear in Section 4.

Eventually, section 5 concludes the paper.

2 Hybrid Retrieval of Visual Content

As mentioned previously, the search engine described hereby supports three modes of queries and retrieval of images and video, namely

1. content-based retrieval,
2. ontology-based retrieval, and
3. hybrid retrieval, which builds upon the combination of the two aforementioned methods.

2.1 Combining visual and semantic information

The main objective behind our retrieval model is to allow a user to complement a query primarily addressed using one of the visual or semantic mode with the other. Starting with one mode, information arising from the complementary mode is used to enhance the results. The additional results presented are considered to be a set of recommendations for the user by broadening the desirable query and they are generated in a transparent way.

The mathematical model of the hybrid retrieval system is described below for both cases. Lets assume that the function $Sem(data_{rdf}, q_{sem})$ is producing the desired output given the data and the query based on the semantic data formed in RDF¹ language by retrieving the results from the Knowledge Base:

$$Res_{sem} = Sem(data_{rdf}, q_{sem}) \quad (1)$$

where $data_{rdf}$ are the metadata stored in RDF in the Knowledge Base and q_{sem} is the query string in RDFQL or SeRQL.

In a similar way, the function: $Vis(data_{desc}, q_{vis})$ outputs the results from content-based search using as data the extracted descriptors of the multimedia content and the proper input from the user.

$$Res_{vis} = Vis(data_{desc}, q_{vis}) \quad (2)$$

where $data_{desc}$ represent the extracted descriptors of the multimedia content and q_{vis} represents the desirable input (i.e one or a set of images) for which visually matching content expected to be retrieved and displayed. The function $Vis(data_{desc}, q_{vis})$ outputs Res_{vis} in a specific ranking based on the similarity coefficient which derives from the calculation of the distances of the extracted descriptors for the objects included in the query.

Subsequently, two cases of hybrid search are defined: (i) the visual search, where the system, given the desirable query, produces visually similar results with the initial object accompanied by a set of recommendations deriving from the transparent semantic query that the visual results produce; and (ii) the semantic search, where a user can submit a query by browsing the ontology fields and acquire the results that illustrate the content which satisfies the constraints of the query complemented by recommendations based on visual similarity of the initial results.

In the case of visual search the output consists of two sets of results: the initial results produced by Res_{vis} and the set of recommendations Rec_{sem} based on semantic search and given by:

$$Rec_{sem} = Sem(data_{rdf}, ResToQ_{sem}(Res_{vis})) \Leftrightarrow \quad (3)$$

$$Rec_{sem} = Sem(data_{rdf}, ResToQ_{sem}(Vis(data_{desc}, q_{vis}))) \quad (4)$$

where the function $ResToQ_{sem}$ creates a new query based on the first set of the results in order to retrieve the semantically related content. The aforementioned function exploits the initial set of results by processing the ontology fields of every output object in order to define the semantic annotation which is mostly shared by these results. The query produced leads to a search for content that shares the specific common value in the chosen ontology fields with the results.

The final set of results Res is the set of results from visual similarity Res_{vis} enhanced by the recommendation results Rec_{sem} :

$$Res = Res_{vis} \cup Rec_{sem}. \quad (5)$$

On the other hand, when a Semantic search occurs the results that are produced consist of: the first set provided by Res_{sem} and the second set Rec_{vis} illustrating the recommendations:

$$Rec_{vis} = Sem(data_{desc}, ResToQ_{vis}(Res_{sem})) \Leftrightarrow \quad (6)$$

$$Rec_{vis} = Vis(data_{desc}, ResToQ_{vis}(Sem(data_{rdf}, q_{sem}))) \quad (7)$$

where the function $ResToQ_{vis}$ constructs a query taking into account the visual features of the initial results. The algorithm used for this function produces descriptors of an average hypothetical object by averaging the descriptors of the results.

The final set of results is:

$$Res = Res_{sem} \cup Rec_{vis}. \quad (8)$$

2.2 Content-based Retrieval

In this retrieval mode, described by (2), users are able to perform a visual-based search by taking advantage of low-

¹<http://www.w3.org/RDF/>

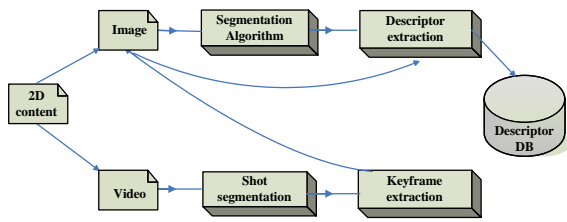


Figure 1. The offline image and video content analysis process.

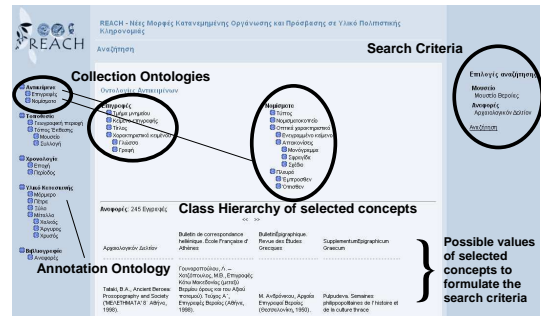


Figure 2. Search engine interface

level multimedia content features. The retrieval system can handle 2D still image and potentially video. In this mode, the user provides, as the input query, an example of the multimedia content she or he is interested in, and, based on the extracted descriptors of the input and the indexed offline-generated descriptors of the content repository, the system performs a visual similarity-based search and the relevant results are retrieved.

Figure 1 illustrates the 2D image analysis with potential application to video files. More specifically, 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 [7]. 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) [8] was applied.

This procedure could be extended or video analysis. In such a case 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.

2.3 Ontology-based Retrieval

This search mode, described by (1), is more appropriate for the cases in which the user knows to an adequate degree of confidence the semantic annotation of the material user provides constraints on the concepts of the five ontologies of Section 3.2. During search time, the system retrieves the semantically connected content according to users selections. The system can handle complex queries that require the combination of multiple concept-based search criteria and thus can retrieve different cultural items that share com-

mon data. Figure 2 illustrates the application of this technique by presenting a proper interface for ontology based search.

3 Search Engine and Evaluation Corpus

The evaluation procedure took place with the employment of a search engine based on the aforementioned retrieval model and it is capable of retrieving cultural visual content. The design of the GUI and the ontology browser plays a significant role for ontology-based retrieval. As shown in Figure 2, the GUI of our search engine provides a view of the ontologies, enabling the browsing through their structure and hierarchy; Apart from the ontology-based retrieval, the search engine supports content-based queries in order to produce results depending on visual similarity. Combining the two techniques, the search engine is capable of providing the hybrid searching functionality as was described in detail in Section 2.1.

3.1 Visual Content

The main content provider is the Center for Greek and Roman Antiquity (KERA)², which offers a large collection of inscriptions and coins, accompanied with detailed documentation. Furthermore, a rich collection of Greek paintings is provided by the Greek museum: Teloglion Foundation of Art³ while a large collection of photographs is offered by Alinari Photographic Archives⁴.

²<http://www.eie.gr/nhrf/institutes/igra/index-en.html>

³<http://web.auth.gr/teloglion/>

⁴<http://www.alinari.com/>

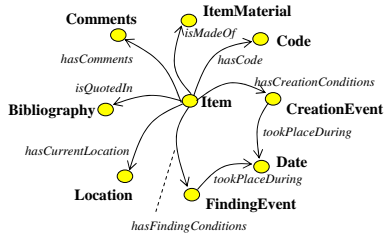


Figure 3. A graphical representation of the concepts in the ontology and their relations.

3.2 Ontology

Cultural heritage collections are accompanied by a rich set of annotations. However, these annotations are 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.

Taking into account the content originally available for our use case scenario an ontology infrastructure has been defined to efficiently describe and represent all knowledge related to each collection. The proposed architecture consists of two layers and makes use of five different RDF(S) ontologies, namely *Annotation*, which is generic, and *Coins*, *Inscriptions*, *Paintings* and *Photographs* which are specific to the collection itemsets of our scenario.

A large set of information fields inside the image annotation set is common for each item, regardless of the collection that is part of. As such, it was decided to use a separate, higher-level ontology specifically intended for representing this kind of concepts and relations. dimensions, etc. Consequently, the role of the *Annotation* ontology (Figure 3) 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 complementary 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.

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 [10] core ontology which has been proposed as an ISO standard for cultural heritage material structuring and representation. To enable this functionality, appropriate mappings between the concepts of our defined ontologies and the CRM were drawn.

3.3 Ground Truth Definition

The ground truth used in order to evaluate the results of the experiments were different for each retrieval mode.

Regarding the content-based experiments as ground truth was considered the (subjective) visual similarity of the objects. More specifically the visual features which were taken into account to prove visual similarity were the shape and the color while the existence of visually related regions between the objects can also be considered as factor of visual resemblance.

The results from ontology-based queries could be easily evaluated due to the existing annotations.

The recommendations, which are results of the hybrid mode, are considered to be results related to the initial set of the semantic or visually based results. The definition of the term recommendation can be subjective however it can be defined as any result which is related to the initial output in terms of visual or semantic similarity.

4 Results

The content-based and ontology-based modes are complementary to each other, and as such, it is meaningless to compare them directly in terms of metrics like precision and recall. In this section the advanced functionalities of the hybrid search engine are demonstrated through use cases and insights into the performance of the two search flavors are provided.

4.1 Hybrid search: use cases

The hybrid search engine is capable of detecting implicit semantic relationships between visually dissimilar images, and extract the relevant artefacts. To demonstrate this capability, two use cases are presented in this section, which are summarized in Figures 4 and 5.

In the first use case (first row of Figure 4), the input query is the painting “Autoprosopgrafia” (“Self-portrait”). During the content-based search visually similar inscriptions are extracted (Figure 4a). As shown in the figure, the results have several visual features in common, such as depiction of human faces and other portraits. To fire an ontology-based search, the system retrieves the most common semantic feature, which happens to be the name of the painter. As such, the system is capable of automatically returning the same results as if the user was capable of submitting a query of the type (in natural language) “Find all the artefacts that look like the painting “Autoprosopgrafia” or all the Paintings created by Pierrakos Alkis” (Figure 4b). The first set of results shows visually similar images by illustrating mostly paintings with human faces and portraits while the recommendations include works of art of the same painter who

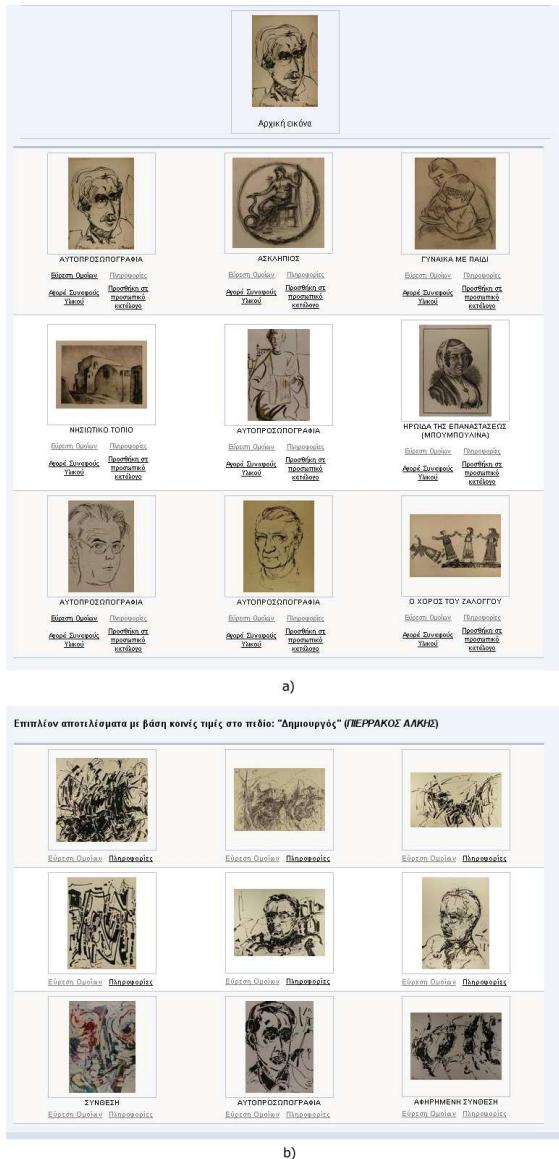


Figure 4. The first use case: (a) initial set of results derived from visual similarity search, (b) set of recommendations based on a complementary semantic query

has created paintings of similar themes (portraits with different technique etc.) complementing in this way the initial results by providing paintings of similar subjects which could not be retrieved by visual similarity due to the different paint technique (Figure 4).

The second use case follows the opposite approach

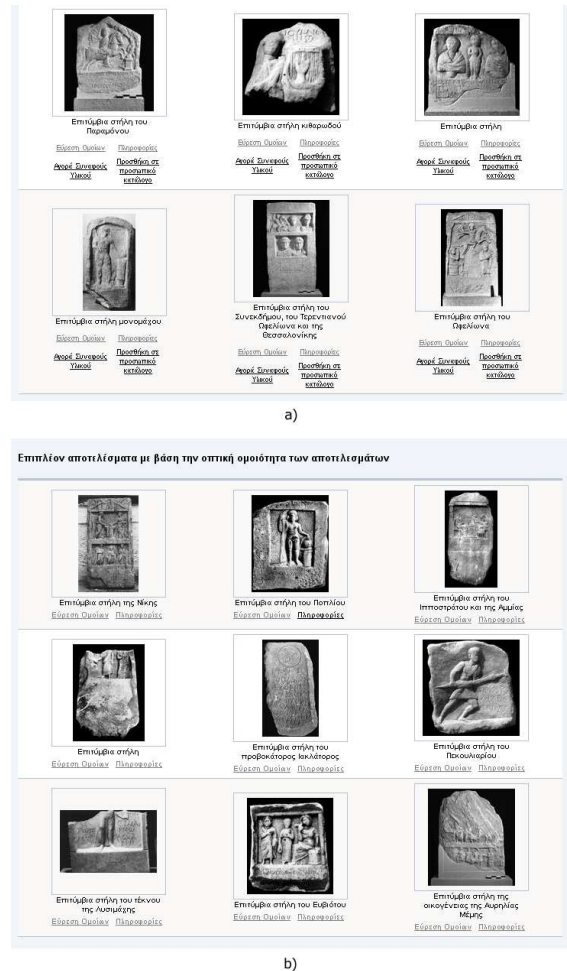


Figure 5. The second use case: (a) initial set of results based on a semantic query, (b) set of recommendations includes visually similar images

where the initial query is based on ontology fields while the set of recommendations derives from visual similarity. In this scenario, the user searches for inscriptions dated in the 4th century BC. The first set of results includes the inscriptions which belongs to the aforementioned period (Figure 5a) while the recommendations provide results visually similar with the initial set of inscriptions. As it can be observed in Figure 5b the recommendation images include more inscriptions with similar shapes and figures with the first set of results. Consequently the recommendations broaden the initial query as they reveal visually similar images of potential interest that the user may not have been aware of when submitted the query.

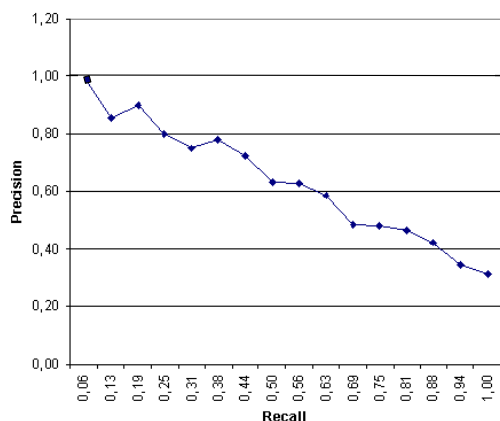


Figure 6. Precision-Recall diagram for the content-based method.

4.2 Performance Insights

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. The dataset is consisted of roughly 4000 images along with a complete set of semantic annotations. The visual descriptors are stored in a collocated MPEG-7 XM server.

Figure 6 shows the Precision-Recall diagram for the content-based retrieval. The curves correspond to the mean precision value that was measured after several retrieval tasks. For the ontology-based search since it is based on selecting available concepts describing the content, the estimation of Precision-Recall diagram is not relevant. The average response time for the ontology-based search is 0.163 sec, while for the content-based search is 0.773 sec.

The behavior of the hybrid search is expected to combine the benefits of the other two approaches providing recommendations to the user in order to broaden the query. Precision-Recall graphs for the recommendations are not presented as these strongly depend on the nature of the retrieval task, and on the objective and purpose of the user when submitting a query.

5 Conclusions

In this paper, a novel retrieval model for handling visual and multimedia digital libraries is presented in an efficient and effective manner. The search engine based on that model adopts three methods for retrieval: two autonomous

and one combinational. The ontology-based method makes use of the formal, logic-based representation of semantic mark-up metadata accompanying each collection, while an illustrative user interface is used for graphical query formulation. Although the search engine dealt with 2D images of cultural heritage content, there is the potential of extension based on the proposed model to include video content. A notable feature of this work is its modular and extensible ontology infrastructure, which provides mappings to CIDOC-CRM in order to gain interoperability with other ontologies from the cultural domain. The hybrid method, which is the main contribution of this work, makes a combined use of the previous two methods. Thus is capable of offering, as a recommendation, a more complete result set to the user, which comprises both visually and semantically similar items, while the input query remains either solely ontology-based or content-based.

References

- [1] E. Riloff and L. Hollaar. Text Databases and Information Retrieval. In *ACM Computing Surveys*, vol. 28, no. 1, pp. 133-135, 1996.
- [2] P. Sinclair, S. Goodall, P. Lewis, K. Martinez, and M. 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.
- [3] T. Gruber. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5:199-220, 1993.
- [4] 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.
- [5] O. Mich, R. Brunelli, and C. Modena. A survey on video indexing. *Journal of Visual Communications and Image Representation*, 10:78-112, 1999.
- [6] E. Izquierdo, J. Casas, R. Leonardi, P. Migliorati, N. O'Connor, I. Kompatsiaris, and M. G. Strintzis. Advanced Content-Based Semantic Scene Analysis and Information Retrieval: The Schema Project. In *Proc. Workshop on Image Analysis For Multimedia Interactive Services, London, UK*, 2003.
- [7] 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.
- [8] MPEG-7 eXperimentation Model http://www.lis.ei.tum.de/research/bv/topics/mmdb/e_mpeg7.html
- [9] V. Kobla, D. Doermann, and K. 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] M. Doerr. The CIDOC-CRM An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3):75-92, Fall 2003.