Improving Semantic Search in Digital Libraries Using Multimedia Analysis

Ilianna Kollia, Yannis Kalantidis, Kostas Rapantzikos, Andreas Stafylopatis
School of Electrical and Computer Engineering, National Technical University of Athens, Zographou Campus 15780, Athens, Greece
Email: ilianna2@mail.ntua.gr, ykalant@image.ntua.gr, rap@image.ece.ntua.gr, andreas@cs.ntua.gr

Abstract— **Semantic search of cultural content is of major importance in current digital libraries, such as in Europeana. Content metadata constitute the main features of cultural items that are analysed, mapped and used to interpret users' queries, so that the most appropriate content is selected and presented to the users. Multimedia, especially visual, analysis, has not been a main component in these developments. This paper presents a new semantic search methodology, including a query answering mechanism which meets the semantics of users' queries and enriches the answers by exploiting appropriate visual features, both local and MPEG-7, through an interweaved knowledge and machine learning based approach. An experimental study is presented, using content from the Europeana digital library, and involving both thematic knowledge and extracted visual features from Europeana images, illustrating the improved performance of the proposed semantic search approach.**

Index Terms— **semantic search, content based search, digital libraries, multimedia analysis, europeana**

I. Introduction

A large activity has been taking place in the field of digital libraries the last few years, taking advantage of new information technologies so as to make cultural heritage more accessible for all. In this framework, probably the most significant achievement has been generation of Europeana[1] as a single reference point for European culture online. Europeana, Europe's digital library, archive, museum and gallery, is probably the most ambitious cultural project ever undertaken at a European scale, bringing together cultural institutions from different sectors and from all the Member States. In the span of less than three years, its collections have grown to more than 22 million digitised objects. Large parts of the digitised works are in the public domain, that is they are no longer covered by intellectual property rights, thus they can in principle be accessed and used for free by everyone.

Cultural institutions add considerable amounts of information (metadata) to digitised objects, describing for example the author, the provenance and age of the work, giving contextual information, as well as technical information on the formats used and characteristics allowing search engines to locate the object. An image of the object, in low resolution, is also attached to its description that is normally good enough for private use, e.g. schoolwork.

The services provided by digital libraries, and Europeana, refer either to content providers, i.e., memory institutions which digitise and place content and metadata online, or to content consumers, such as researchers, students, general users, enabling annotation, discussion and user-generated content. Technologies and related services which have received much attention in the field are Interoperability and Semantic Search [15], [19], [20]. Interoperability may require transformation of metadata to common standards, such as the Europeana Semantic Element (ESE) and Europeana Data Model (EDM), or creation of mappings between metadata used by different content providers. Our 'MINT' platform[2] has been used in more than ten Europeana projects for mapping and aggregating for Europeana more than 4 millions items; it is currently studied by the evolving Digital Public Library of America[3] for mappings among heterogeneous sources of cultural content.

Semantic search targets on answering user queries, by exploiting both explicit and implicit related knowledge. Reasoning on available knowledge bases [16], based on appropriate representations and languages, such as description logics, RDF, Web Ontology Language (OWL) [1], [5], [6] can be the means to move ahead in this direction. The creation of linked data stores [22], [23] from digital cultural heritage resources enables the linking of multiple data, assisting efficiency by permitting combined or linked searches.

Nevertheless, the usage of multimedia information, and particularly of the - free from property rights - provided images, for improving the search results has not been investigated so far, apart from some aside developments [21].

In this paper, we exploit both semantic metadata representations and images included in a Digital Library such as Europeana, using state-of-the-art knowledge-based reasoning and multimedia analysis. We show that improved query answering can be obtained if we interweave semantic technologies with machine learning paradigms [11] applied on appropriate features extracted from the images of the Digital Library.

A diagram of the proposed approach focusing on the semantic based search and its interweaving with content-

[1] http://www.europeana.eu

[2] http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki

[3] http://blogs.law.harvard.edu/dpla/

doi:10.4304/jmm.7.2.193-204

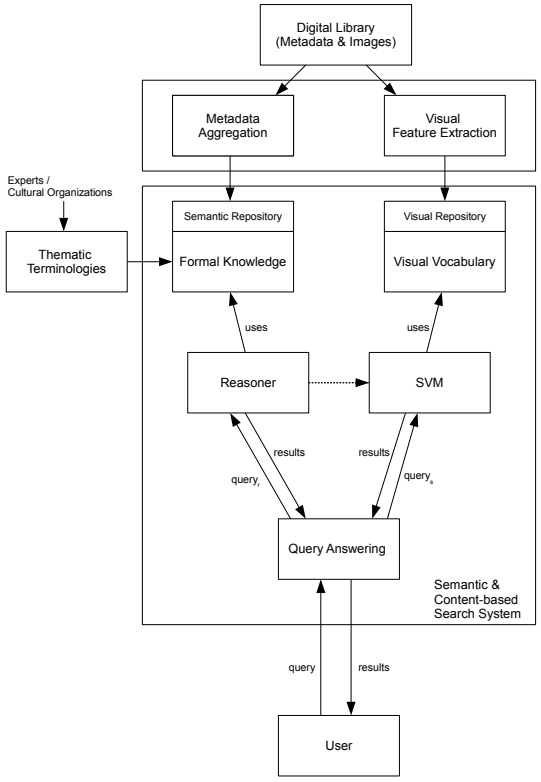


Figure 1. The architecture of the proposed query answering system

based search for providing effective answers to users' queries is presented and analysed in Section II. The extraction of visual features, the design of the visual vocabulary and their usage for describing the Digital Library items are described in Section III. Section IV presents a scheme in which the available knowledge used in the semantic search is interweaved with machine learning, the latter operating on the feature sets extracted from the cultural images. An evaluation study illustrating the theoretical developments is presented in Section V. Conclusions and future work are given in Section VI.

II. Semantic and Content-based Search in Digital Libraries

A. The proposed framework

Figure 1 presents a block diagram of the proposed system which is able to semantically analyse users' queries and provide answers exploiting both the metadata and the respective visual representations of the Digital Library objects.

The Digital Library may consist of a single portal, as in the case of Europeana, or of a network of content providers, as will probably be the case of the evolving Digital Public library of America. In the case of Europeana, interoperability during users' searches is achieved by developing and using a common metadata model (Europeana Data Model) that is compatible with the Dublin Core or METS standard used by libraries, as well as with LIDO that is used by museums, EAD used by archives and EBU Core used by audiovisual archives. The metadata model elements are descriptions of the objects providing basic and advanced information about them, starting from the answers to 'Who?', 'What?', 'When?' and 'Where?'. Moreover, the location (url) of the image of each digital object is normally given together with a low resolution version of it. When users submit a query for an object, the respective metadata of the objects are searched and whenever a match is achieved the object is included in the results returned to the user, using some answer ranking scheme[4]. In our system, the metadata are aggregated and represented as RDF triples (forming the formal assertional knowledge) in terms of the EDM ontology. They are then stored in the *Semantic Repository*.

If, however, we want to let users ask complex queries and receive appropriate answers, we need a more detailed description of cultural content in the form of terminological knowledge in various domains (*Thematic Terminologies*). In this paper, we show that, whenever such knowledge is available, we can develop semantic search and semantic query answering, i.e., construct answers to queries posed by users, based not only on string matching over the digital library metadata, but also on the implicit meaning that can be extracted by reasoning using the terminological knowledge [2], [3], providing details about species, categories, properties, interrelations (e.g., brooches are made of copper or gold). The latter knowledge (the Thematic Terminologies) is developed by experts or the content providers, and is stored in the *Formal Knowledge Subsystem*. The metadata elements in the Semantic Repository are represented as descriptions of individuals, i.e., connections of individuals with entities of the terminological knowledge.

The creation of global axioms that hold over all items of the digital library is, however, very difficult. One approach to deal with this is to use only axioms containing constraints that are known to hold over all data and leave out of the formal knowledge any constraint that holds over most (but not all) of the data. In either case, the inherent, or resulting, incompleteness of the formal knowledge poses limitations to its usage in semantically searching and answering queries over cultural heritage content.

In the proposed approach we show that this problem can be partially overcome if we also use content-based search of the Digital Library images. Content-based search over the images included in the digital library can provide another source of results to users' queries. This has not been exploited so far in the development of digital libraries, such as Europeana, which have been evolving by extending the metadata-based description, search and access methodology used by libraries, archives and museums so far. In Figure 1, an image feature extraction and visual repository generation stage runs in parallel with metadata aggregation. Appropriate visual vocabularies are

[4]http://www.europeana.eu

then generated and exploited by machine learning techniques; the target being to improve the semantic search of cultural heritage content. Effective extraction of visual features and generation of visual vocabularies is presented in the next section.

B. The Semantic Search Algorithm

The representation formalism used for the terminological descriptions is OWL 2 (the W3C Standard for Ontology representation on the web) [6]. The theoretical framework underpinning the OWL 2 ontology representation language is *Description Logics* (DL) [1]. The DL language underpinning OWL 2 is $\mathcal{SROIQ}$.

The building blocks of DL knowledge bases are atomic concepts, atomic roles and individuals that are elements of the denumerable, disjoint sets $\mathbf{C}, \mathbf{R}, \mathbf{I}$, respectively. A DL knowledge base (KB) is denoted by $\mathcal{K}=\langle\mathcal{T},\mathcal{A}\rangle$, where $\mathcal{T}$ is the terminology (usually called TBox) representing the entities of the domain and $\mathcal{A}$ is the assertional knowledge (usually called ABox) describing the objects of the world in terms of the above entities. Formally, $\mathcal{T}$ is a set of terminological axioms of the form $C_1 \sqsubseteq C_2$ or $R_1 \sqsubseteq R_2$, where C_1, C_2 are $\mathcal{SROIQ}$-concept descriptions and R_1, R_2 are $\mathcal{SROIQ}$-role descriptions. $\mathcal{SROIQ}$-concept expressivity employs conjunction ($C_1 \sqcap C_2$), disjunction ($C_1 \sqcup C_2$), universal and existential quantification ($\forall R.C$, $\exists R.C$), qualified number restrictions ($\geq R.C$, $\leq R.C$) and nominals ($\{a\}$), while $\mathcal{SROIQ}$-role expressivity allows for the definition of role inverse (R^-) and role compositions ($R_1 \circ R_2$) in the left part of the role inclusion axioms. $\mathcal{T}$ describes the restrictions of the modeled domain. The ABox $\mathcal{A}$ is a finite set of *assertions* of the form $A(a)$ or $R(a,b)$, where $a,b \in \mathbf{I}$, $A \in \mathbf{C}$ and $R \in \mathbf{R}$.

An interpretation $\mathcal{I}$ maps concepts to subsets of the object domain, roles to pairs of elements from the object domain and individuals to elements of the object domain. For an interpretation to be a model of a knowledge base several conditions have to be satisfied [7]. If an axiom ax is satisfied in every model of a knowledge base $\mathcal{K}$ we say that $\mathcal{K}$ entails ax, written $\mathcal{K} \models ax$. Entailment checks are performed by appropriate tools called *Reasoners*.

We next consider concept queries. A concept query q is of the form $q : Q(x) \leftarrow \bigwedge_{i=1}^{n} C_i(x)$, where x is a variable and $C_i(x)$ are predicates-concept atoms. An example of a concept query is $Q(x) \leftarrow OpenVase(x) \wedge VaseWithTwoHandles(x)$ consisting of two concept atoms. An individual a is an answer/instance of a concept query q posed over the DL knowledge base $\mathcal{K}$ iff $\mathcal{K} \models Q(a)$. The procedure we follow to find the answers to concept queries is shown in Algorithm 1

The algorithm takes as input the system's formal knowledge base $\mathcal{K}$ and a user's query q and returns the results, i.e., the individuals of the knowledge base that satisfy the query. This is done by iterating over the concept atoms C_j of the query q and over the individuals a appearing in the knowledge base $\mathcal{K}$ and by checking whether $\mathcal{K}$ entails that a is an instance of C_j. If the instantiated concept atom is entailed we add the individual to the set of answers *Ans* else, if it is not, we have to check whether a is already contained in *Ans*. In this case we remove it from the set or else we leave the set as it is.

Algorithm 1 Query Evaluation Procedure using the high level knowledge

Input: $\mathcal{K}$ $\langle\mathcal{T},\mathcal{A}\rangle$: the $\mathcal{SROIQ}$ knowledge base DL ontology
q: a concept query
Output: Ans: the set of answers to query q

```
1: Ans := ∅
2: C^1,...,C^n:=queryAtomsOf(q)
3: for j=1, ..., n do
4:     for all individual a ∈ A do
5:         if K ⊨ C_j(a) then
6:             Ans := Ans ∪ a
7:         else
8:             if a ∈ Ans then
9:                 Ans := Ans \ a
10:            end if
11:        end if
12:    end for
13: end for
14: return Ans
```

III. Image Feature Extraction and Visual Vocabulary Generation

A. MPEG-7 Visual Feature Extraction

The need for extracting descriptors in a standardized way has led to the MPEG-7 standard [31] that focuses on the description of multimedia documents. It is composed of a set of audio, color, texture and shape descriptors that have been used for image classification, high-level concept detection and image/video retrieval.

The visual descriptors, which have been considered in our analysis are color, shape and texture ones, including the Dominant Color Descriptor (DCD - a set of dominant colors in a region or in the image), the Color Structure Descriptor (CSD - capturing both global color features of the image and local color spatial structure), the Color Layout Descriptor (CLD - a resolution-invariant visual descriptor designed to represent the spatial distribution of color in the YCbCr color space), the Scalable Color Descriptor (SCD - a Haar-transform based transformation applied across values of the image's color histogram), the Region-Based Shape descriptor (RSD - expressing the 2-D pixel distribution within an object or a region of interest based both on the contour and inner pixels), the Homogeneous Texture Descriptor (HTD - a quantitative characterization of texture based on oriented filters) and the Edge Histogram Descriptor (EHD - capturing the edges' spatial distribution).

Figure 2 presents some images from Europeana, which are among the answers to a user query asking for 'Jewelry of the Late Byzantine period'; they include crosses and mosaics. The corresponding segmentation masks are also

shown, produced by thresholding and median filtering. MPEG-7 shape descriptors have been extracted from the segmented images. Distance values of the Region Shape descriptor for similar and non-similar images (of Figure 2) are shown in the confusion matrix of Table I. It can be easily seen that Distance is high for non-similar images and low for similar ones. This permits usage of an SVM classifier to discern the two different categories. Moreover, if a user makes a search within a study on religion, objects that are known in the knowledge base to be related to religion, such as crosses, should be first searched for, followed, if necessary, by more searches in a scalable way. In this case, if an image of a cross, such as the ones shown in Figure 2 is selected due to its thematic annotation, then the two other crosses shown in the same Figure can also be selected, irrespectively of whether their thematic annotation is complete or not, since their feature sets match that of the already selected image.

Figure 3 shows the Color Structure and Homogeneous Texture descriptor histograms for three of the images shown in Figure 2. Two of them refer to similar images (562, 558) with the last (3535) being different. It is evident that color and texture histograms can also be used to distinguish between the different categories. Figure 4 presents images of items belonging to the 'brooch' category, that are made either of gold or copper. The binary masks of the foreground objects are also computed and used to extract color descriptors from the corresponding regions, while discarding the background. Items made of copper share similar color distributions, with these distributions being different from the item made of gold, as shown in Figure 5. Based on this, an SVM classifier can separate brooches such as the ones shown in Figure 4 (h,i), that are made of copper, from brooches as the one in Figure 4 (g) that are made of gold. It should be mentioned that, in fact, the annotation of the images shown in Figure 2 (a,b) do not include the material which they are made of, while the annotation of the, very similar, image shown in Figure 2 (c) indicates that it is made of the material lazourite. Categorization of the images can, therefore, be used for content enrichment, where incomplete annotation exists, leading to improved search and answers to users' queries.

TABLE I.
CONFUSION MATRIX (L1 DISTANCE) OF REGION SHAPE MPEG-7 DESCRIPTOR

image id	(e)	(f)	(a)	(b)
(e)	0	85	181	171
(f)	85	0	154	148
(a)	181	154	0	120
(b)	171	148	120	0

B. Local Feature Extraction and Visual Vocabulary Generation

Among the most popular local features and descriptors are the Affine-covariant regions [48], i.e., regions that follow the affine transformations of the underlying image structure and are robust to occlusion and viewpoint changes. According to the early study by Mikolajczyk *et al.* [39], the best performing detectors have been the Maximally Stable Extremal Region (MSER) [38] and Hessian-affine [39] detectors.

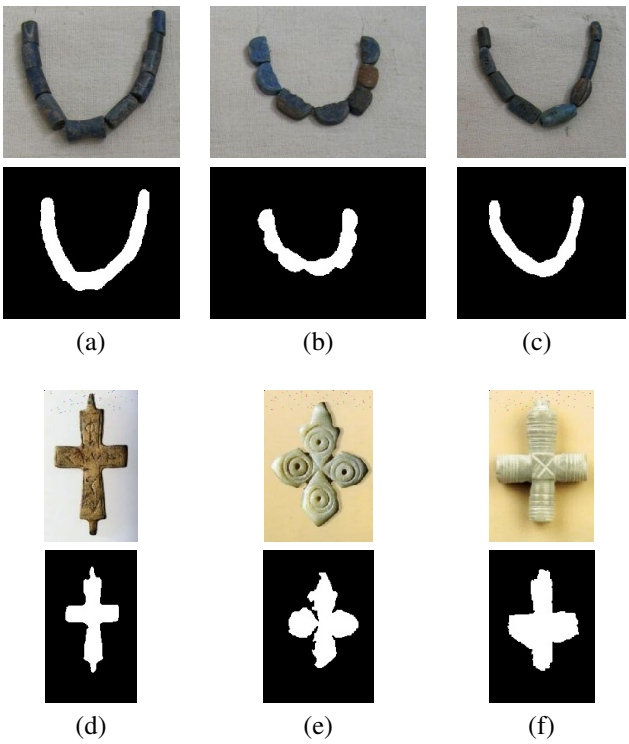


Figure 2. Sample images and corresponding segmentation masks.

The need to tackle the trade-off between computational complexity and performance led to a variety of local feature detectors mainly based on intensity/gradient distribution. Hence, in an attempt to promote precision, Rapantzikos *et al.* proposed an edge-based detector [45] that detects composite regions from single-scale edges. Furthermore, Avrithis *et al.* [28] proposed the Medial Feature Detector (MFD), which is based on a weighted distance map on image gradient. On the side of computational efficiency, several detectors have been proposed, like CenSurE [50], FAST [53], BRIEF [49], with SURF (Speeded-Up Robust Features) [29] being still the most popular one. The SURF detector is based on integral images for fast image convolutions (scale-space) and a resulting approximation of the Hessian matrix. Since large-scale visual similarity with small computational complexity is the aim of the proposed in this paper system, we use SURF features and descriptors for the representation of the Digital Library images.

An example of the extracted SURF features is shown in Figure 6, where sample images of vessels and the associated features are depicted. Local features are detected on the foreground object only, where corner-like and blob-like structures are present.

Towards robust image representation from local features, computer vision researchers have adopted the *Bag-of-Words* (BoW) model, or -synonymously- *visual vocabulary* as an equivalent to a typical language vocabulary,

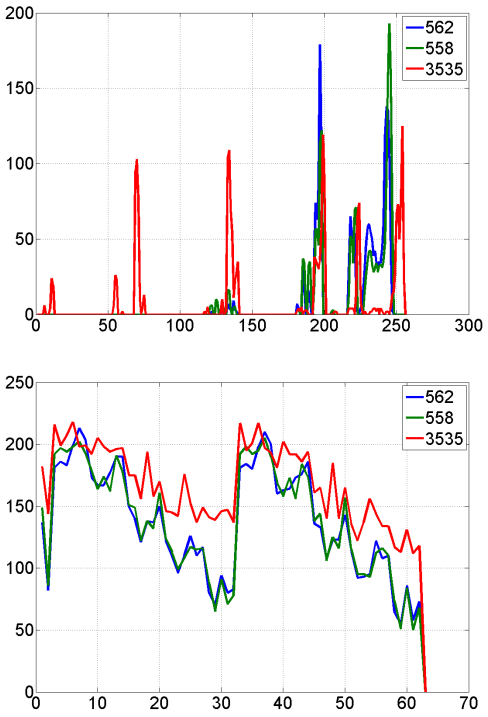


Figure 3. Histogram comparisons for MPEG descriptors (CST: top, HT: bottom.). Two histograms correspond to two similar images, while the third one is an irrelevant one. (better shown in color)

Figure 4. Sample images and corresponding segmentation masks. g: made of gold, h-i: made of copper.

with an image corresponding to a text document. Descriptors are extracted on local features and are quantized in order to build the vocabulary, which is then used to describe each image by the set of visual words it contains [47], [32]. We adopt this model for describing the cultural heritage images.

To obtain the visual vocabulary for representing the image, we create clusters in the space of descriptors and assign each feature to the closest centroid (i.e., visual word). We should note here that due to their polysemy, visual words cannot be as accurate as natural language words. This means that a given visual word cannot be directly assigned to a concept, but it can represent a part of a significantly large number of concepts. Generally, these errors can be adequately compensated for by employing

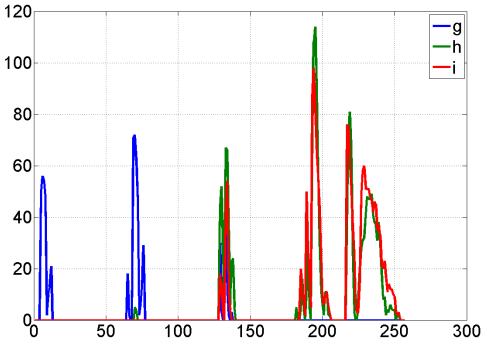


Figure 5. Histogram comparison for CST MPEG descriptor extracted from the regions defined by the binary masks of Figure 4.

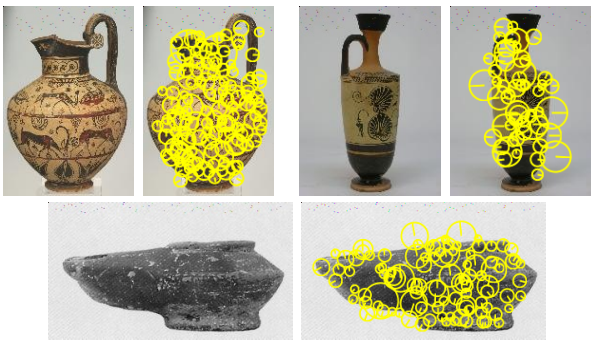


Figure 6. Sample vessel images and corresponding SURF local features extracted. Local shape of features is shown in yellow circles with scale and dominant orientation.

a spatial verification stage of matched features.

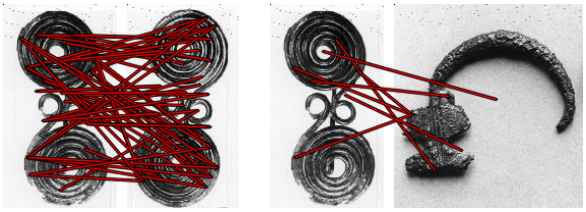
Typically, the visual vocabulary creation is performed using the k-means clustering algorithm [37]. However, if the number of the points to be clustered is significantly large, clustering using the k-means algorithm becomes a very slow task. We use a fast variant of the k-means algorithm that uses approximate nearest neighbor search, *i.e.* nearest cluster centers at each iteration are assigned using randomized kd-trees [46]. Specifically, we use the FLANN library of Muja and Lowe [42] both in vocabulary creation and in assigning visual words to image features.

Having each local feature assigned to a visual word, we can represent each image in terms of the visual words it contains. A histogram of constant-length can be constructed for each image, containing the appearance frequencies of each visual word. This is the BoW histogram, a N_v-dimensional vector $H_v(I)$ of an image I:

$$H_v(I) = [tf_I(0), tf_I(1), \ldots, tf_I(N_v)] \quad (1)$$

where $tf_I(i)$ denotes the number of times that the visual word i was selected as a nearest neighbor of one of the interest points extracted from image I.

An example with cultural heritage images is shown in Figure 7. Pairwise image similarity is higher when many common visual words appear in the images and the histogram bins have similar values, i.e., in similar images



(a) (b)

Figure 7. Local features correspondences based on visual word assignments. (a) Matching features between two similar images. Similar images share many common visual words. (b) Matching features between two non-similar images

and lower in non-similar ones.

Vector $H_v(I)$ constitutes the first part of the input provided to the SVM in Figure 1; the second part comes from the above-mentioned MPEG-7 features.

IV. Interweaving Semantic Search with Machine Learning of Visual Features

In Figure 1 we use machine learning techniques, in particular support vector machines, to learn from the extracted visual features to classify the digital Library items in various concepts that can appear in queries. As a consequence, these techniques can determine the items that satisfy the corresponding query concept atoms, irrespectively of whether these items have been identified by the Reasoner that exploits the formal knowledge of the system. This results in bridging the gap between restrictions imposed by ontologies and actual restrictions (visual features) that each cultural heritage item possesses.

Support Vector Machines (SVMs) constitute a well known method which is based on kernel functions to efficiently induce classifiers that work by mapping the visual features, and the corresponding items, onto an embedding space, where they can be discriminated by means of a linear classifier. As such, they can be used for effectively exploiting the extracted features and classify the cultural heritage items in the different concept categories that are included in the formal knowledge. We use one SVM for each concept. The input vector of the SVM includes $N1+N2$ elements, N1 of which are the SURF of the visual vocabulary and N2 the selected MPEG-7 features. Each type of features is appropriately normalized. Hence, MPEG-7 descriptors are normalized using L_1 or L_2 norm (depending on their type [31]) and SURF using the L_1 norm. Finally, all features are normalized according to L_1 norm and are fed to the SVM.

In the proposed approach the visual based search aims at improving the accuracy of the semantic search. For this reason we use the successful responses provided by the reasoner exploiting the formal knowledge as training samples for each SVM, which is then used to test its performance to new inputs. This is shown by a dotted arrow from the reasoner towards the SVM in Figure 1.

The kernel used in the SVM to encode the visual knowledge through similarity between different images is a normalized linear kernel defined as follows [9]:

$$k_l(x,y) := \frac{x^T y + c}{\|x\|\|y\|} \tag{2}$$

where x,y are vectors of features, $\|\cdot\|$ is the Euclidean norm and c is considered zero.

The *Query Answering Subsystem* merges the outputs of the two modules, i.e, the reasoner and the SVM module.

It is, however, possible to extend the SVM kernel so as to include the knowledge referring to the individuals that contribute to the Formal Knowledge Ontologies [12], [14] and consequently let only the output of the SVM module feed the Query Answering subsystem.

The extension comes from a family of kernel functions defined as $k_p^F : Ind(A) \times Ind(A) \to [0,1]$, for a knowledge base $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$. $Ind(A)$ indicates the set of individuals appearing in A, and $F = \{F_1, F_2, \ldots, F_m\}$ is a set of concept descriptions. These functions are defined as the L_p mean of the, say m, simple concept kernel functions κ_i , $i = 1,\ldots,m$, where, for every two individuals a,b, and $p > 0$,

$$\kappa_i(a,b) = \begin{cases} 1 & (F_i(a) \in A \wedge F_i(b) \in A) \vee \\ & (\neg F_i(a) \in A \wedge \neg F_i(b) \in A); \\ 0 & (F_i(a) \in A \wedge \neg F_i(b) \in A) \vee \\ & (\neg F_i(a) \in A \wedge F_i(b) \in A); \\ \frac{1}{2} & \text{otherwise.} \end{cases} \tag{3}$$

$$\forall a,b \in Ind(A) \quad k_p^F(a,b) := \left[\sum_{i=1}^{m} \left| \frac{\kappa_i(a,b)}{m} \right|^p \right]^{1/p} \tag{4}$$

The above kernel encodes the formal knowledge for the problem under analysis through the similarity of pairs of individuals with respect to high level features, i.e. concepts of the knowledge base.

The rationale of these kernels is that similarity between items is determined by their similarity with respect to each concept F_i , i.e., if two items are instances of the concept or of its negation. Because of the OpenWorld Assumption for the underlying semantics, a possible uncertainty in concept membership is represented by an intermediate value of the kernel. A value of p = 1 is generally used for implementing (4).

The extension we can use is a combined SVM kernel, computed as the mean value of the above described two kernels, i.e., $k_c(a,b) = k_p^F(a,b) + k_l(a,b)$ where k_p^F is the above knowledge driven kernel and k_l is the normalized linear kernel.

Let us now sketch the way that queries are evaluated using SVMs that have already been trained to classify cultural heritage items to concepts.

Algorithm 2 shows the procedure. The algorithm takes as input the data we want to query together with their visual features and uses trained SVMs to check which items simultaneously belong to all concepts appearing in the query. $SVMpredict$ predicts the label of an item w.r.t. a concept C_j using the SVM trained to classify items to this concept.

Algorithm 2 Query Evaluation Procedure using the low level visual features

Input: $trainedSVM$: a vector of trained SVM
Input: $data$: the queried data
Input: $features$: the features of the $data$
q: a concept query
Output: Ans: the set of answers to query q

```
Ans := ∅
C^1,...,C^n:=queryAtomsOf(q)
for j=1, ..., n do
    for all a ∈ data do
output(a) := SVMpredict(features(a), trainedSVM_j)
        if output(a) = 1 then
            Ans := Ans ∪ a
        else
            if output(a) = 0 then
                Ans := Ans \ a
            end if
        end if
    end for
end for
return Ans
```

Algorithm 3 Combined Query Evaluation Procedure

Input: KB : the $\mathcal{SROIQ}$ knowledge base
Input: SVM : the vector of trained SVM
Input: $data$: the queried data
q: a concept query q
Output: Ans: the set of answers to query q

```
Ans := ∅
Ans_KB := Evaluate_KB(KB, q)
features := ExtractFeatures(data)
Ans_SVM := Evaluate_SVM(SVM, data, features, q)
Ans := Ans_KB ∪ Ans_SVM
return Ans
```

Algorithm 3 shows how we interweave the two approaches, the knowledge based and the kernel based approach to enrich in this way the search results. The method $extractFeatures$ extracts the features of the images, while methods $evaluate_{KB}$ and $evaluate_{SVM}$ perform Algorithms 1 and 2 to extract the query answers from the two methods which are then disjuncted and given to the user.

V. Evaluation Study : Improving Semantic Search through Visual Analysis

A. Setting Up the Experimental Study

The experimental study presented in this section aims at illustrating the improvement which is achieved by exploiting both semantic and content-based search of a Digital Library, while answering users' queries. We focus on Europeana, because it is a real environment, easily accessible by everyone through the Europeana portal.[5] The content we consider is, on the one hand, Hellenic content, consisting of about 40,000 items and, on the other hand, museum content aggregated by the ATHENA project [27], consisting of about 3,600,000 Europeana objects. The metadata of this content was transformed to EDM OWL. The experiments presented are based on the Hellenic content, since for this content we also possess thematic knowledge; we make reference to the ATHENA content when dealing with scalability issues.

TABLE II.
Excerpt of the used thematic ontology in description logic syntax

$Bowl \sqsubseteq OpenVase$
$Beetle \sqsubseteq OpenVase$
$Cup \sqsubseteq OpenVase$
$Crater \sqsubseteq OpenVase$
$Basin \sqcup Lekanis \sqsubseteq OpenVase$
$Crate \sqsubseteq OpenVase$
$Crate \neq Crater$
$Amphora \sqsubseteq VaseWithTwoHandles$
$Beetle \sqsubseteq VaseWithTwoHandles$
$Bowl \sqsubseteq VaseWthTwoHandles$
$Crater \sqsubseteq VaseWithTwoHandles$
$Pelike \sqsubseteq VaseWithTwoHandles$
$Amphora \sqsubseteq BigVase \sqcap CloseVase$
$Alabaster \sqsubseteq VaseWithoutHandles$
$Amphora \neq Alabaster$
$CloseVase \neq OpenVase$
$VaseWithTwoHandles \neq VaseWithOneHandle$
$VaseWithTwoHandles \neq VaseWithoutHandles$

The thematic knowledge we used for Hellenic monuments, particularly for vases (for which metadata and images are provided) has been created in the framework of the Polemon and 'Digitalisation of the Collections of Movable Monuments of the Hellenic Ministry of Culture' Projects of the Directorate of the National Archive of Monuments[6] and which has been included in the Polydefkis terminology Thesaurus of Archaeological Collections and Monuments [25], [26]. This knowledge contains axioms about vases in ancient Greece, i.e., class hierarchy axioms referring to the different types of vases, such as amphora, alabaster, crater, as well as axioms regarding the appearance, usage, creation period and the material vases were made of. It contains 55 categories of cultural objects (such as pottery, jewelry, stamps, wall paintings, engravings, coins) and more than 300 types. An excerpt of the used terminological knowledge referring to types of vases is shown in Table II.

Based on this knowledge and the analysis of the related images, our target is to answer user queries asking, e.g., for Pottery of Mycenaean period, Minoan pottery with sea pace decoration, Jewellery of Hellenistic period, Molyvdovoula (king's stamps) of the Middle and Late Byzantine period, Coins of the Late Byzantine period, Open vases, Vases with two handles, Open vases with two handles.

As a consequence, the TBox of the formal knowledge of our system consists of the EDM ontology together with the thematic ontology. The ABox consists of the EDM instances (Europeana items) each one of which is described by its type, its creation date, its material, the

[5]http://www.europeana.eu

[6]http://nam.culture.gr

museum it can be found at. About 1,000,000 RDF triples have been generated from the 40,000 Europeana items and stored in a Sesame repository.

Apart from the metadata, the visual features of the cultural objects were extracted according to the methodology presented in Section III. Implementing the methodology for SURF feature extraction and visual vocabulary generation requires tuning of some parameters used in it. Specifically, the approximate number of features (tuned through a threshold) has been selected to be 500, while the number of visual words has been set as a tuning parameter, varying in the range of 100-4,000, according to the related literature. Regarding the MPEG-7 features, either both color and texture features were used, providing a vector with 666 elements, or only the texture (color) features were used, providing a vector with 142 (524) elements respectively.

B. The Query Answering Results

In the following we applied the Query Answering approach shown in Figure 1, focusing on the techniques described in Sections II and IV for semantic query answering and its interweaving with the extracted visual information on the above mentioned Europeana items. In our implementation we used the HermiT reasoner [3], [4] and the LIBSVM library[7] for reasoning over the knowledge base and for learning the visual descriptors respectively.

Let us consider three of the user queries mentioned in the former subsection, in particular the ones shown in the first column of Table III, i.e., queries for 'vases with two handles', or 'open vases', or 'open vases with two handles'.

The second column reports the accuracy of the query answering task, defined as the number of true positives and true negatives over all test data, when the Algorithm 1 of Section II is used. The knowledge base that this algorithm takes as input includes both the terminological knowledge described in the previous subsection and the Europeana instance data. About 4,000 items out of the 40,000 ones belonged to the concept Vase, with about 2,200 being open vases, 1,800 being vases with two handles and with around 900 belonging to both categories. Ground truth for all these items has been created by experts from the cultural sector.

It can be easily verified that in all three queries, the knowledge used for the definition of open vases and/or vases with two handles accounted for about 85% of the cases, resulting in an error from 10 to 15%.

Let us now consider the usage of content-based search as a means to improve the above-derived query answering accuracy. In particular, we first train SVMs (one for each of the two different concepts of the queries) using the extracted visual features of the items which are returned as query answers by the knowledge base. In particular, the SVMs are trained using the normalized linear kernel

[7]http://www.csie.ntu.edu.tw/~cjlin/libsvm/

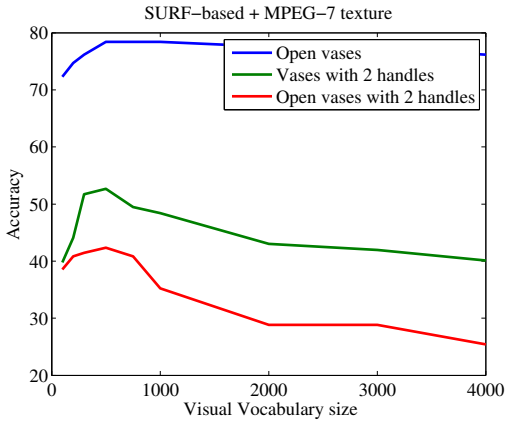


Figure 8. Obtained Accuracy with respect to the size of visual vocabulary

described in Section IV, based on the visual features and the annotated labels of the images, for all items that have been correctly returned from the reasoner used above in Algorithm 1. In all three cases the SVMs learned to classify correctly all provided training data.

Various combinations of the visual inputs mentioned above were considered as inputs to the SVMs: a) only the MPEG-7 features, resulting in an input vector with 666 elements, b) only the SURF-based visual vocabulary, with the number of visual words varying between 100 and 4,000, c) all MPEG-7 and visual worlds, d) a combination of the texture MPEG-7 features (with 142 elements) and visual words, that was identified to provide the best results.

Following training, we tested - according to Algorithm 2 - the performance of the SVMs to the remaining data (about 10-15% in all three cases), which, erroneously, have not been returned as query answers based on the knowledge.

We made a variety of experiments with the cases (a)-(d) mentioned above and analysed the performance of the SVMs in terms of: i) accuracy, ii) precision (defined as the ratio of true positives over the returned, by the SVM, answers) and iii) recall (defined as the ratio of true positives over the theoretical correct answer).

Figure 8 presents the obtained accuracy in all three queries for the above-mentioned (d) case, when varying the number of visual words, which show that the best results were obtained using a vocabulary of 500 visual words. Accuracy is higher in the open vase category and lower in the two-handle vase one. Figure 9 shows the corresponding precision and recall values for a vocabulary size between 100 and 4,000 words, which verify the results derived from Figure 8. For this number of visual words (500), Table IV shows a comparison of the performance obtained by each one of the (a)-(d) test experiments, verifying that the combination of visual words and MPEG-7 texture features provides the best results as measured by all criteria.

The third column of Table III summarises the accuracy of query answering when we use Algorithm 2 of Section

TABLE III.
ACCURACY (%) OF QUERY ANSWERING

Query	Accuracy(Algorithm 1)	Accuracy(Algorithm 2)	Accuracy(Algorithm 3)
1. $Q(x) \leftarrow OpenVase(x)$	85.5	78.4	96.9
2. $Q(x) \leftarrow VaseWithTwoHandles(x)$	84.6	52.7	92.1
3. $Q(x) \leftarrow OpenVase(x) \wedge VaseWithTwoHandles(x)$	90.2	42.4	93.2

TABLE IV.
COMPARISON OF SVM PERFORMANCE FOR DIFFERENT VISUAL INPUT VECTORS

Visual Features	Accuracy			Precision			Recall		
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
SURF-based only	55.68	51.61	42.37	93.48	71.79	100	54.43	43.08	26.15
MPEG-7 only	67.05	48.39	30.51	93.1	71.79	100	68.35	43.08	14.58
SURF-based + MPEG-7	64.77	49.46	30.51	92.86	72.5	100	65.82	41.54	14.58
SURF-based + MPEG-7 texture	78.41	52.69	42.37	96.88	73.68	100	78.48	43.08	25

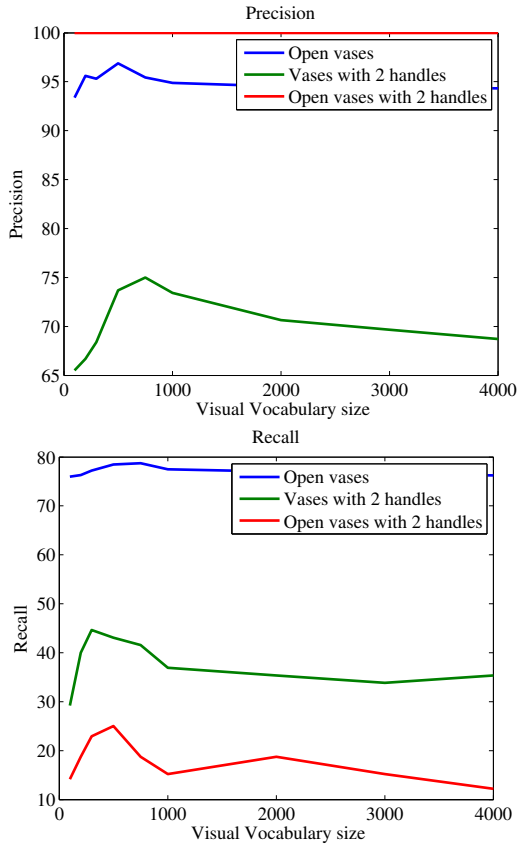


Figure 9. Obtained Precision and Recall in each of the three queries

IV. The above means that column 3 shows the percentage of the data that the SVMs 'correctly' predicted as query answers among those that were not predicted as such by the knowledge.

The fourth column of Table III refers to the accuracy of query answering when Algorithm 3 is used to combine the results of the knowledge based and visual kernel based approaches. We see that the accuracy of the algorithm, computed as the number of generated query answers that are true has been increased in all cases. This illustrates the improved performance of the semantic search when visual information is taken into account based on the procedures of Sections II, III and IV.

For example, vases with two handles are defined in the knowledge base to comprise the categories of amphora, beetle, bowl, crater, pelike. Nevertheless, there are also some vases in other categories, such as in jug or basin, which may have two handles. These items will not be among the results that knowledge based query answering (Algorithm 1) will return to the user. It is in such cases, that the interweaving with visual feature based machine learning (Algorithms 2 and 3) improves the performance of the semantic search service offered to the users. In Table III, the former method achieves a performance of 85.5%, 84.6% and 90.2% in the queries. The proposed combined knowledge and machine learning approach rises the performance to 96.9% , 92.1% and 93.2% respectively.

C. Scalability Issues: A Visual Search Engine for Europeana (VIEU)

An issue that deserves special attention in the implementation of the proposed approach is scalability.

It can be shown that query answering over OWL 2 knowledge bases is computationally intensive, suffering from high worst-case complexity. This is certainly a serious issue, if we want to apply the proposed method to a large Digital Library, such as Europeana, where the number of triples will be in the order of millions or billions. The methodology we can follow in order to face this problem is: a) to use the ability of the triple store system to provide approximate answers based on the materialisation method [8], thus reducing the computational load, b) to take advantage of the highly modular form of the metadata and of the terminologies, so as to partition them in a set of much smaller independent knowledge bases. This modular character is mainly a result of the different metadata origination (libraries, archives, museums) and the respective thematic diversity.

Let us also focus on the visual feature extraction, visual vocabulary generation and linking of different images based on their visual characteristics. As shown in the former subsection, implementing the procedure for feature extraction and visual vocabulary generation requires tuning of some parameters used in it. Specifically, the approximate number of features (tuned through a threshold) and the number of visual words should be

defined. To be able to use the proposed procedure in the large scale environment of a Digital Library, we performed a large-scale retrieval experiment based solely on images included in the ATHENA content provided to Europeana. The developed tool, namely the Visual Search Engine for Europeana, can be evaluated on-line at http://vieu.image.ntua.gr, including about 10% of the ATHENA items' images, i.e., around 375,000 images retrieved from the Europeana portal.

Local feature extraction, visual vocabulary construction and the assignment of visual words have been performed in an offline step for all the images. Following this, visual representation of all images has been organized in an index structure. The visual vocabulary has been extracted, using a set of ~10,000 images and its size has been chosen after a trial-and-test procedure, during which the best performing vocabulary has been selected. In particular, for this database size we selected a visual vocabulary of 30,000 clusters (with the number of local features being ~500 per image). Top-ranked images were passed to a re-ranking stage where they have been checked for spatial verification. Finally, spatial matching is performed with Fast Spatial Matching (FSM) [51] and/or Hough Pyramid Matching (HPM) [52].

The outcome of this procedure, which the VIEU system offers, is the ability to obtain links among as many items' images as possible based on their visual similarity. This can assist and speed up the content-based query answering proposed in this paper, especially if linked through the extended SVM kernel with the available formal knowledge, when dealing with large Digital Library sizes. Moreover, the VIEU system can be extended using tags in a straightforward way. Similarly to the Visual Image Retrieval and Localization (VIRaL) system found in http://viral.image.ntua.gr, tags related to visually similar images can be transferred to the query image. These tags are usually aggregated from the images' source site or enriched by exploiting links to external sites, e.g., the Wikipedia.

VI. Conclusions

The current state-of- the-art in Digital Libraries, including Europeana, targets towards effective access to content, taking advantage of semantic technologies, such as knowledge representations, metadata standards and mappings, linked data and user modeling. Searching for information over all content repositories and semantically processing the associated content has been one of the main goals in this evolution. Content metadata constitute the main feature of cultural items that are analysed, mapped and used to interpret users' queries, so that the most appropriate content is presented to them. In this framework the research and development focus is on generating effective and efficient advanced search mechanisms. Taking advantage of multimedia information, mainly visual, that is offered to the users in digital library records, has not gained much attention so far (especially in formation of the Europeana portal). This is due to problems, on the one hand, with intellectual property rights of the associated cultural objects and, on the other hand, with the inherent difficulty of analysing images, especially in the usually offered low resolution version of them. The current paper presents a new semantic search methodology, including a query answering mechanism that can meet the semantics of users' queries and at the same time enriches these answers by exploiting appropriate visual features, both local and MPEG-7, through an extended knowledge and machine learning based approach. We have shown that by visual feature extraction, clustering, machine learning and interweaving with knowledge driven answering to users' queries, improved semantic search can be achieved in digital libraries. In July 2012 Europeana fully adopts the CC0 Creative Commons Model for Intellectual Property Rights regarding content metadata. This will permit free usage of the Europeana content metadata and items' images. Our plan is to extend the generated VIEU system so as to include all Europeana images, thus facilitating access and content-based queries to the whole Europeana content. Moreover, applying the improved semantic search in a variety of subareas, such as archaeology, photography, modern arts, fashion, where specific, and modular, thematic knowledge can be derived and used, as well as combining it with the evolving field of linked open data for cultural heritage are future extensions of the presented work.

Acknowledgment

The authors wish to thank the Hellenic Ministry of Culture and Tourism for their assistance in working with the cultural content of the www.collections.culture.gr.

References

[1] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press (2007)

[2] Perez-Urbina, H., Motik, B., Horrocks, I.: Tractable query answering and rewriting under description logic constraints. Journal of Applied Logic 8(2), 186-209 (2010)

[3] Ilianna Kollia, Birte Glimm and Ian Horrocks, SPARQL Query Answering over OWL Ontologies. In: Proceedings of the 8th Extended Semantic Web Conference (ESWC 2011). Springer Verlag (2011).

[4] Rob Shearer, Boris Motik and Rob Shearer and Ian Horrocks, HermiT: A Highly-Efficient OWL Reasoner, Proc. of the 5th Int. Workshop on OWL: Experiences and Directions (OWLED 2008 EU) (2008)

[5] Frank Manola and Eric Miller, editors. *Resource Description Framework (RDF): Primer*. W3C Recommendation (2004)

[6] Boris Motik, Peter F. Patel-Schneider and Bijan Parsia, editors. *OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax*. W3C Recommendation (2009)

[7] B. Motik, P.F. Patel-Schneider, B. Guenca Grau (eds.): OWL 2 Web Ontology Language Direct Semantics. W3C Recommendation (27 October 2009), available at http://www.w3.org/TR/owl2-direct-semantics/

[8] H. J. Horst. Completeness, decidability and complexity of entailment for RDF Schema and a semantic extension involving the OWL vocabulary. *Journal of Web Semantics*, 3(2-3):79-115, (2005)

[9] Bernhard Schoelkopf and Alexander. J. Smola. Learning with Kernels. The MIT Press, Cambridge, Massachusetts, 2002

[10] Richard Duda, Peter Hart, David Stork. Pattern Classification. John Wiley & Sons, Inc, 2001

[11] Simon Haykin. Neural Networks A Comprehensive Foundation. Prentice Hall International, Inc., 1999

[12] Ilianna Kollia, Nikolaos Simou, Andreas Stafylopatis and Stefanos Kollias. Semantic Image Analysis using a Symbolic Neural Architecture. Journal of Image Analysis and Stereology.2010.

[13] N. Fanizzi, C. d'Amato, F. Esposito. Learning to Rank Individuals in Description Logics Using Kernel Perceptrons. RR 2010, 4th International Conference on Web Reasoning and Rule Systems, September, 22 - 24, 2010, Bressanone/Brixen, Italy.

[14] N. Fanizzi, C. d'Amato, F. Esposito. Statistical Learning for Inductive Query Answering on OWL Ontologies. ISWC 2008, International Semantic Web Conference, October, 26-30, 2008, Karlsruhe, Germany.

[15] SIEDL: First Workshop on Semantic Interoperability in the European Digital Library, 5th European Semantic Web Conference, Tenerife, Spain, June 2, 2008.

[16] Hitzler, P., Kroetzsch, M., Rundolph, S.: Foundations of Semantic Web Technologies. Chapman & Hall/CRC (2009)

[17] Report 'Existing standards applied by European Museums', http://www.athenaeurope.org

[18] Report 'Converting XML files to Apenet EAD', http://www.apenet.eu

[19] The New Renaissance Report of the European Reflection Group on Digital Libraries (Comite des Sages), January 10, 2011, http://www.europeana.eu

[20] Report of European Commission Member State Expert Group on Digitisation and Digital Preservation (MSEG), http://www.ec.europa.eu

[21] http://ec.europa.eu/information_society/apps/projects/

[22] Christian Bizer, Tom Heath, Tim Berners-Lee, Linked Data: The Story So Far, International Journal on Semantic Web and Information Systems (IJSWIS), 2009.

[23] Keynote speech "Linked Data and Europeana: Perspectives and issues", Europeana Plenary Conference, The Hague, The Netherlands, September 14, 2009.

[24] Report of Joint Programming Initiative in Cultural Heritage and Global Change, http://www.cordis.europa.eu/pub/fp7/coordination

[25] M. Doer, D. Kalomirakis, "A Metastructure for Thesauri in Archaeology, Computing Archaeology for Understanding the Past". In Proceedings of the 28th Conference, BAR International Series, Lublijana, 2000.

[26] D. Kalomirakis, "Polydefkis: A Terminology Thesauri for Monuments". In M. Tsipopoulou (ed.), Proc. of "Digital Heritage in the New Knowledge Environment: Shared spaces and open paths to cultural content", Athens, 2008.

[27] Proceedings of the ATHENA Conference "Cultural Institutions Online", Rome, 28 April 2011.

[28] Y. Avrithis and K. Rapantzikos. Capturing boundary structure for feature detection. In *International Conference on Computer Vision (ICCV)*, 2011.

[29] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In *ECCV*, 2006.

[30] J.L. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.

[31] Shih-Fu Chang, Thomas Sikora, and Atum Puri. Overview of the MPEG-7 Standard. *IEEE trans. on Circuits and Systems for Video Technology*, 11(6):688–695, 2001.

[32] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, volume 1, page 22. Citeseer, 2004.

[33] J.H. Freidman, J.L. Bentley, and R.A. Finkel. An Algorithm for Finding Best Matches in Logarithmic Expected Time. *ACM Transactions on Mathematical Software (TOMS)*, 3(3):209–226, 1977.

[34] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, 1988.

[35] T. Lindeberg. Feature detection with automatic scale selection. *IJCV*, 30(2):79–116, 1998.

[36] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[37] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 1–297, 1967.

[38] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[39] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.

[40] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.V Gool. A comparison of affine region detectors. *IJCV*, 65(1):43–72, 2005.

[41] A.W. Moore. An intoductory tutorial on kd-trees. Technical report, Technical Report.

[42] Marius Muja and David G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *International Conference on Computer Vision Theory and Application VISSAPP'09)*, pages 331–340. INSTICC Press, 2009.

[43] A. Neubeck and L. Van Gool. Efficient Non-Maximum Suppression. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 3, 2006.

[44] S.M. Omohundro. Efficient algorithms with neural network behavior. *Complex Systems*, 1(2):273–347, 1987.

[45] K. Rapantzikos, Y. Avrithis, and S. Kollias. Detecting regions from single scale edges. In *International Workshop on Sign, Gesture and Activity (SGA'10), European Conference on Computer Vision (ECCV)*, 2010.

[46] C. Silpa-Anan and R. Hartley. Optimised kd-trees for fast image descriptor matching. 2008.

[47] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *International Conference on Computer Vision*, volume 2, pages 1470–1477, 2003.

[48] T. Tuytelaars and L.J Van Gool. Content-based image retrieval based on local affinely invariant regions. In *Visual Information Systems*, 1999.

[49] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *European Conference on Computer Vision*, ser. Lecture Notes in Computer Science, K. Daniilidis, P. Maragos, and N. Paragios, Eds., vol. 6314. Springer, 2010, pp. 778–792.

[50] M. Agrawal, K. Konolige, and M. R. Blas, "CenSurE: Center surround extremas for realtime feature detection and matching," in *ECCV*, 2008.

[51] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *CVPR*, 2007.

[52] G. Tolias and Y. Avrithis, "Speeded-up, relaxed spatial matching," in *In Proceedings of International Conference on Computer Vision (ICCV 2011)*, Barcelona, Spain, November 2011.

[53] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *European Conference on Computer Vision*, vol. 3951. Springer, 2006, p. 430.

Ilianna Kollia was born in Athens in 1986. She completed her Diploma in Computer Science in the Electrical & Computer Engineering (ECE) School of the National Technical University of Athens (NTUA) in 2009 and the M.Sc. in Computer Science in the Department of Computer Science of the University of Oxford, UK in 2010. Since 2010 she is a Doctorate student in the ECE School of NTUA. She is currently collaborating with the University of Ulm, Germany , using a DAAD short term research scholarship. Her research interests include knowledge representation, reasoning and adaptation, query answering, multimedia and cultural heritage applications.

Yannis Kalantidis is a PhD candidate at School of Electrical and Computer Engineering of National Technical University of Athens. He is currently conducting research in the area of web-scale image search, clustering and localization. He is a graduate of the School of Electrical and Computer Engineering of the National Technical Univarsity of Athens since February 2009.

Dr. Konstantinos Rapantzikos received the diploma the MS degree in Electronic and Computer Engineering from the Technical University of Crete (Greece) in 2000 and 2002 respectively, and the Ph.D. degree from NTUA in 2008. His interests include computational modeling of human vision, visual action recognition, biomedical imaging and motion estimation. He has published 12 articles in international journals and books and 25 in proceedings of international conferences.

Prof. Andreas-Georgios Stafylopatis received the Diploma degree in Electrical and Electronics Engineering in 1979 from the National Technical University of Athens and the Docteur IngÃl'nieur degree in Computer Science in 1982 from the UniversitÃl' de Paris-Sud, Orsay, France. Since 1984 he is with the School of Electrical and Computer Engineering at the National Technical University of Athens, where he is currently a professor of computer science leading the Intelligent Systems Laboratory. His research interests include computational intelligence, machine learning, data mining, knowledge discovery, multimedia and web applications, intelligent agents and recommendation systems.