A knowledge-based image retrieval system integrating semantic and visual features

Olfa Allani\textsuperscript{a,b,}\textsuperscript{*}, Hajer Baazaoui Zghal\textsuperscript{a}, Nedra Mellouli\textsuperscript{b,c}, Herman Akdag\textsuperscript{b}

\textsuperscript{a}RIADI Laboratory, ENSI, University of Manouba, Tunisia
\textsuperscript{b}LIASD Laboratory, Paris 8 University, Saint-Denis, France
\textsuperscript{c}IUT of Montreuil, Paris 8 University, Saint-Denis, France

Abstract
The main limitations of the existing high level image retrieval approaches concern the high dependance on an external reliable resource (domain ontologies, learning sets, etc.) and a model for mapping semantic and visual information.
In this paper, we propose an image retrieval system integrating semantic and visual features. The idea is to automatically build a modular ontology for semantic information and organize visual features in a graph-based model. Both elements are then combined together in a same component called "pattern" used for retrieval. The system has been implemented and the obtained results show that our proposal enables an improvement in the retrieval task.

Keywords: image retrieval; region; ontology module

1. Introduction

Different image retrieval approaches combining semantic and visual aspects of images have been proposed. There are approaches that aim to expand the level of information by extracting a semantic level given the visual information; approaches based on users feedback (relevance feedback) and approaches bringing together visual and semantic aspects into advanced models.
The approaches aiming to expand the information level use automatic and semi-automatic semantic image annotation techniques to annotate images for further use (retrieval, storage etc.). The semi-automatic techniques rely on user interaction which is essential to provide a basic annotation support\textsuperscript{17}, whereas the automatic ones are based on learning semantic knowledge from a large number of images and then using the concept models to annotate new images. Approaches based on user’s feedback have been proposed to deal with the difficulty and cost of providing rich and reliable textual annotations for images. In this new line of reasoning, users are included in the retrieval loop in order

\textsuperscript{*}E-mail address: olfa.allani@gmail.com
to identify what they are looking for\textsuperscript{4}. The application of relevance feedback algorithms has shown a great interest in the retrieval process\textsuperscript{15}. Finally approaches based on image and text combination aim to provide a rich knowledge representation that improves machine interpretation\textsuperscript{13,10}. They are mainly inspired by ontology-based Information retrieval works.

Despite the large amount of works, the problem of mapping between semantic and visual image aspects is still a challenging issue in image retrieval field.

In this paper, we propose an image retrieval system integrating semantic and visual features. The idea is to automatically build a modular ontology based on textual corpus enriched by a terminological resource and to organize visual information in a graph-based model. The resulting module and graph are combined together to represent a unique component called "pattern". Thus, our main contribution concerns the integration of the visual and semantic information in a given pattern.

The remainder of this paper is organized as follows. Section 2 presents an overview of the related research, along with our motivations and objectives. Section 3 describes our system architecture and its different components. In section 4, a case study illustrating the functionality of the proposed system as well as the application results are presented. Finally, we conclude this paper by summarizing our proposal and presenting some future research perspectives.

2. Related search, motivations and objectives

A wide range of works have been conducted aiming to reduce the semantic gap by modeling high level semantics including visual facets. The idea underlying these approaches is to provide a model for mapping visual information to high level semantics.

Since images are usually provided with external meta-data, many approaches exploiting these meta-data for image retrieval have been proposed. They can be classified in three main categories, namely Natural Language Processing (NLP) based, Description Logic (DL)-based and Ontology-based approaches. The aim is to transform the meta-data into a format that can be naturally used for image retrieval.

For example, NLP-based approaches use natural language analysis tools such as OpenNLP\textsuperscript{1}, RelEx\textsuperscript{6}, Stanford nlp\textsuperscript{9}, and social media analytic tools, to extract relevant semantic information from images context. The DL-based approaches rely on description language formalization of external data. They propose logic-based models\textsuperscript{16} to formalize the semantic dimension of multimedia content using a logical formalism such as the description logic which is considered as the most important knowledge representation formalism. Finally, the ontology-based approaches use external data resources available or especially built for this aim aside from images, to carry on the retrieval process. These approaches can be characterized, given their generality level, as domain dependent or general approaches. For example, we distinguish object ontology models and multi-modality ontologies etc. Object ontology models\textsuperscript{14} require a taxonomic and lexical processing of concepts. Multi-modality ontologies include semantic concepts and relations as well as visual classes of images resulting from image classification and relationships extracted from low-level features\textsuperscript{10}.

The introduced approaches focused on providing mechanisms for mapping low-level features to specific semantic concepts in order to reduce the semantic gap. However, this problem is still partially resolved since the proposed approaches strongly depend on an external reliable resource such as large learning sets, powerful formal languages or valid domain ontologies. In particular, the key problem of ontology-based approaches is the integration of a domain ontology built independently from the image database semantic content. Moreover, the diversity of domains presented in a same image database makes it complex to describe, with a single ontology, the overall database semantics. Furthermore, to reduce this complexity, one possible solution is to use modular ontologies, since modularization has become a key solution in ontology engineering\textsuperscript{5}.

To deal with these limitations, we propose a system that has two main objectives: On the one hand, modeling a deeply semantic structure that covers the different domains and maintains all of the possible connections using modular ontologies; on the other hand, introducing a novel image index denoted as "pattern". This allows an accurate mapping between low level features and high level modular ontology-based features. Low-level features describe

\textsuperscript{1} https://opennlp.apache.org/
image regions contents, while high level modular ontology features describe the semantic contents implicitly related to image regions. The relations between patterns are also formalized as edges of the graph of patterns which reflect relations between the ontology modules.

The originality of our proposal concerns the integration of modular ontologies to cover different domains within a given image database. The main challenges are: first, to design a meaningful, coherent, minimally redundant and efficiently detailed modular ontology; second, to combine visual features and modular ontology advantages to improve the image retrieval process.

The main steps of the proposal are: (1) Building automatically, on the overall image domains, a modular ontology that integrates autonomous and coherent ontology modules; (2) Building automatically a visual graph of image’s regions for each domain vocabulary. These graphs are fully connected and edges represent visual similarity between regions; then integrating patterns into a graph of patterns. The graph is composed of nodes representing patterns and edges representing semantic and taxonomic relations.

3. An Image Retrieval system based on semantic and visual features

In this section, we describe the system’s architecture and detail the different image indexation and retrieval phases and steps. The general process is performed on two phases; an online phase and an offline phase. We point out that the input is an annotated and segmented image database. The offline phase is supported by two main components namely

data pre-processing (cf. Figure 1 (1)), and pattern graph building (cf. Figure 1 (2)). First, a domain extraction to determine the key domains covered by the database is needed (cf. Figure 1 : 1.1). Then, for each domain, a textual corpus is built given the annotations of the images (cf. Figure 1 : 1.3). A domain-based region’s repartition step is
also needed to prepare for the region’s graph building (cf. Figure 1:2.2). Second Pattern graph building component consists of 4 steps: ontology module building, region’s graph building, pattern building and pattern graph building. Ontology module building is based on domain vocabulary as well as an external resource (cf. Figure 1:2.1). To build the region’s graph, we adopt the SIFT feature extraction algorithm8 to get feature vectors associated to each region. A SIFT-based similarity is computed to get the final region’s graph of a domain (cf. Figure 1:2.2). Both ontology module and region’s graph are integrated to get a pattern (cf. Figure 1:2.3). The final step is the pattern’s graph building which aims to establish relations between resulting patterns using the semantic and taxonomic relations previously extracted (cf. Figure 1:2.4). The reason for which we use a graph formalism is to organize and provide a rich mapping between semantic and visual information in order to narrow down the semantic gap between the low-level visual features and the richness of human semantics. The online phase starts when the user submits a query to the system. A query could be an annotated or non-annotated image. This phase is composed of three main steps. First, the user query is segmented and SIFT features are extracted (cf. Figure 1:segmentation). For annotated queries, a textual extraction of the annotation’s concepts is also performed. Then, similarity measures are computed in order to select candidate similar patterns (cf. Figure 1:similarity). When similar patterns are selected, an advanced search within the patterns is performed in order to retrieve similar images. The resulting images are then ranked and displayed to the user.

3.1. Data pre-processing component

As depicted in figure 2 the main steps of the data pre-processing component are namely domain extraction, domain-based regions, corpus building and concepts and relations extraction. The data pre-processing is structured around concepts clustering and domains extraction. As mentioned above, the image database covers several domains (eg. vehicles, animals, landscape, insects and music instruments, etc.). The concepts/words clustering consists on grouping words which are semantically similar (having a close specific meaning) words together. Several algorithms have been applied to this aim. In our case, concept clustering consists on partitioning a set of concepts based on their semantic distances into homogeneous groups. The output is a set of clusters representing different domains (cf. Figure 1 (II)). To achieve this goal, we adapt the clustering algorithm proposed in2, which consists on assigning concepts according to maximization the simwup distance. Based on the obtained concepts, the data pre-processing component offers meanings for a given concept from the terminological resource and a selection of the adequate concepts based on a semantic disambiguation algorithm11. The terminological resource choice is primordial since we seek a complete, well organized and scalable resource to be adapted to our needs. Thus, we selected BabelNet.
which is a multilingual encyclopedic dictionary and a semantic network automatically integrating 12 of the mostly used resources. Indeed, if the referent of the target concept has many senses according to BabelNet (Babel synset), the semantic disambiguation task is made using Babelfy.

After synset identification, we generate from a domain a set of sub-domains based on the similarities between each couple (concept, synset) in the domain (cf. figure 2). Each resulting vocabulary is the origin of a textual corpus that is extracted from images annotations (cf. figure 2). The output of this step is a set of domains as well as their associated textual corpuses. A set of regions described with concepts of this domains vocabulary is also created. The corpuses obtained from the previous component is the input of the next component which aims to build patterns.

3.2. Pattern graph building component

Pattern building component consists of two main steps, namely, ontology module building and regions graph building. An ontology module is an ontology fragment that includes only the concepts and relations relevant for a domain. It describes the semantics of data giving a common understanding of the domain and therefore provides a vocabulary enrichment through their different relations. We point out that each ontology module is characterized by a key concept, called pivotal concept that encapsulates the basic meaning of the granule of knowledge.

First, to extract semantic relations, the text corpus is segmented into sentences. Depending on the sentence grammatical gender (with or without verb), a relation extraction algorithm is applied. Indeed, semantic relations are extracted from sentences with verbs using existing tools. Whereas, they are extracted from non-verbal sentences, based on an algorithm that we proposed as a complement. It consists of detecting, using Natural Language Processing (NLP) tools, the predicates that represent semantic relations such as spatial, temporal and causality relations.

Once semantic relations extraction is carried out, filtering is then done in order to get the subset of semantic relations involving the concepts related to the domain. The vocabulary enrichment is performed. Using BabelNet, synonyms, hyperonyms and hyponyms are considered. Relationships can be classified into hyponymy or specialization relationships generally known as "is kind of" or "is-a" and partitive relationships or meronymy which describe concepts that are parts of other concepts.

The resulting sets of taxonomic and semantic relationships as well as the resulting set of concepts are the basis of the formalization. The formalization consists of transforming the resulting vocabulary into a knowledge representation language dedicated to create ontologies.

The next step is the region’s graph building. It represents the visual facet of the pattern and is composed of nodes (regions) and edges (similarity measures). A region graph is a fully connected and non-oriented graph relating the regions representing a same domain vocabulary. Each edge is quantified with a visual similarity measure defined on the SIFT feature vector extracted from the corresponding region. Once visual similarities among the regions of the graph are computed, they are sorted to select delegate regions (with highest similarity values). The SIFT feature represents an image content descriptor. The distances between these complex feature vectors cannot be measured using Euclidean metric or other well known metrics. Thus we proposed a SIFT-based similarity metric.

The ontology module and the regions graph are the two components integrated together to compose a pattern. Therefore, a pattern can be defined as a composition of visual and semantic primitives that appears frequently in image data and that convey rich information mapped to the high level semantic content. A pattern is identified using two elements which are the pivotal concept of the ontology module and the set of delegate regions. Finally, the pattern’s graph is a graph where nodes represent the patterns and edges represent taxonomic/semantic relations established between ontology modules of the patterns.

3.3. Pattern-based image retrieval: online phase

The online phase includes three main steps which are: query segmentation into regions, features extraction, similarity computing based on pattern graph (cf. figure 1 Online phase). First, a query image is introduced by the user, it is segmented into regions using a unsupervised learning algorithm. Then, visual features are extracted (SIFT feature). The output is a feature vector extracted from each query image region. In this work, queries have been manually

---

2 http://babelnet.org/
segmented in the context of another work whose purpose was the evaluation of segmentation algorithms. Finally, the retrieval of semantically similar images is performed on each region of the query. Figure 3 illustrates the retrieval process.

**Fig. 3. Pattern-based retrieval**

4. Case study and Experimental results

Our proposal has been implemented and evaluated in order to show its interest in image retrieval. From more technical point of view, we implemented a prototype using Java programming language. The source codes of the SIFT visual feature implementation were provided by the LIRe library.

4.1. Experimental setup

To evaluate our approach, we used the Image CLEF 2008 and the SIAPR-TC 12 data-set. The SIAPR-TC 12 data-set is composed of 20,000 annotated images in different sizes with a JPEG extension. Images are heterogeneous and represent very diverse themes. English annotations are provided. Images are clustered into regions labeled with a vocabulary of 267 words. They are also indexed with various visual features as SIFT, circularity, etc. There are 117 query images (with the qrels file) in different sizes (average image dimensions: 480*360 pixels) and representing various objects and themes.

In order to measure the image retrieval effectiveness, we used as evaluation metrics:

- The exact precision measure (P@5, P@10, P@15, P@20, P@30 and P@100)
- The exact recall measure (R@5, R@10, R@15, R@20, R@30 and R@100)

In order to evaluate the results improvement and further investigate the prototype performance level, we define two different image retrieval strategies which are: Visual pattern-based strategy (VP) and Visual then semantic pattern-based strategy (SVP).

4.2. Case study

Throughout this section, we illustrate a case study of the system’s application in image retrieval. We start with the offline phase steps, particularly, pattern graph building. We highlight the pattern building steps with a domain example. Then, we introduce the online phase proceeding protocol as well as the retrieval results of one of defined strategies for a query image.

The data pre-processing consists on domain extraction using concepts clustering. Our clustering algorithm generates 46 concepts’ classes. The taxonomic relations extracted from the IAPR-TC benchmark hierarchy represent the relations responsible for establishing inter-pattern links. We extracted 78 hierarchical relations between concepts related to the different domains. The next step consists in building the associated domains patterns. The domain of "mountains" has been selected to present he different pattern building steps. Our domain’s basic vocabulary is represented as

---

3 https://code.google.com/p/lire/
4 http://imageclef.org/SIAPRdata
a set of concepts. Domain vocabulary = \{"mountain", "volcano", "cliff", "hill", "ridge", "alp"\} where the pivotal concept is "mountain". Then, corpus building is performed by collecting annotations of images containing one of the concepts of the domain vocabulary. The next step consists in extracting related concepts, taxonomic and semantic relations. Using BabelNet enrichment, we obtained a large set of taxonomically and semantically related concepts. Some of the obtained relations and concepts are presented in table 2.

### Table 1. An excerpt of semantic and taxonomic relations

<table>
<thead>
<tr>
<th>Relation</th>
<th>Type</th>
<th>Concept 1</th>
<th>Concept 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_A</td>
<td>Taxonomic</td>
<td>volcano</td>
<td>mountain</td>
</tr>
<tr>
<td>CLIMBING</td>
<td>Semantic</td>
<td>climber</td>
<td>mountain</td>
</tr>
<tr>
<td>LIVING_IN</td>
<td>Semantic</td>
<td>person</td>
<td>mountain</td>
</tr>
</tbody>
</table>

The visual aspect is represented with the graph of regions. For the mountain concept, we collected 8107 regions. We computed SIFT-based similarities for each two regions associated to a same concept in the mountain domain. The resulting graph is composed of a set of sub graphs, each associated to a concept from the domain. The edges of the graph are the SIFT-based similarity measures. At the end of this step we select a number of representing regions that are considered as the patterns visual identifiers (delegate regions).

The online phase, which consists on image retrieval based on the semantic and visual pattern-based is illustrated in figure 4 (b). The relevant results are represented in green frames. We distinguish seven relevant results on the top 30 results, distributed as follows: three relevant images on the top 10, then five on the top 20. Figure 4 (a) illustrates the query image which represents a volcanic landscape, it is segmented into regions representing the sky, the volcano
and some trees. The obtained results represent visually close images however, the difference lies in the fact that the mountains images are considered as non relevant results.

4.3. Experimental results and discussion

In order to evaluate the impact of high-level information integration into indexation step, we propose a set of experiments based on visual queries. We propose three retrieval strategies based on patterns as well as a classical content based retrieval denoted SCTBIR (Shape, Color and Texture Based Image Retrieval)\(^1\). SCTBIR uses the canny edge detector, the color layout and the edge histogram to retrieve similar images. It is obvious that the pattern-based retrieval outperforms others by returning more relevant images with higher ranking.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>P@10</th>
<th>P@20</th>
<th>P@30</th>
<th>P@100</th>
<th>Improvement @20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.1425</td>
<td>0.1328</td>
<td>0.1309</td>
<td>0.1291</td>
<td>-</td>
</tr>
<tr>
<td>New Baseline</td>
<td>0.1782</td>
<td>0.1654</td>
<td>0.1502</td>
<td>0.1317</td>
<td>-</td>
</tr>
<tr>
<td>VP</td>
<td>0.1853</td>
<td>0.1754</td>
<td>0.1698</td>
<td>0.1685</td>
<td>32.08% - 27.94%</td>
</tr>
<tr>
<td>SVP</td>
<td>0.1818</td>
<td>0.1742</td>
<td>0.1592</td>
<td>0.1574</td>
<td>31.17% - 19.51%</td>
</tr>
</tbody>
</table>

The strategies VP and SVP use the SIFT feature to retrieve images based on the pattern graph. Compared to a classical content-based strategy using color and texture features, we remarked that our strategies outperforms them. Especially, VP leads an increase in the exact precision at low recall (P@10 and P@20).

In order to judge the proposed system, we compare our system to related works\(^7,10\). Compared to\(^7\), our system automatically performs the semantic enrichment. On the contrary of the works\(^10\), we propose a generic approach where we automatically build ontology modules for different domains. Throughout the conducted experiments we noticed that the two phases pattern-based retrieval provides better retrieval results in terms of precision and recall. They also bring a semantic meaning to retrieved results since information about selected patterns is also retrieved.

5. Conclusion

This paper describes our system for pattern-based image indexation and retrieval. Our goal is to organize the image database in a graph of patterns automatically built for the different domains covered by the image database. The pattern building relies on visual graphs establishment and automatic ontology modules building. Our system takes place in 3 main phases which are: automatically building, on the overall image domains, a modular ontology that integrates autonomous and coherent ontology modules, building automatically a visual graph of image’s regions for each domain vocabulary and building the graph of patterns on the whole image database. The graph is composed of nodes representing patterns and edges representing semantic and taxonomic relations. We implemented our system based on the pattern components and developed our own ontology module building algorithm. The experiments that have been carried out highlight an improvement in retrieved results compared to other content-based image retrieval approaches. Indeed, it contributes to significantly increase the relevance of retrieval results, by enhancing the ranking of images. This improvement confirms our proposal about pattern’s role in semantic and visual information enhancement ans thus, in semantic gap reduction.

In future works, we intend to expand the perimeters of tests to a larger image database such as ImageNet in order to evaluate system’s behavior. We also intend to improve our segmentation algorithm and compare our results with manually segmented images.
References


