Towards Visual Vocabulary and Ontology-based Image Retrieval System

Jalila Filali [1], Hajer Baazaoui Zghal[1] and Jean Martinet[2]
[1]*RIADI Laboratory , National School of Computer Science, University of Manouba, Tunis, Tunisia*
[2]*CRIStAL Laboratory, University of Lille 1, Lille, France*

Keywords: Image Retrieval, Visual Vocabulary, Ontologies.

Abstract: Several approaches have been introduced in image retrieval field. However, many limitations, such as the semantic gap, still exist. As our motivation is to improve image retrieval accuracy, this paper presents an image retrieval system based on visual vocabulary and ontology. We propose, for every query image, to build visual vocabulary and ontology based on images annotations. Image retrieval process is performed by integrating both visual and semantic features and similarities.

1 INTRODUCTION

Several approaches have been introduced and applied to image retrieval. In this context, two basic image retrieval approaches have been proposed in literature: 1) content based image retrieval (CBIR) and 2) semantic image indexing and retrieval (SIIR). In CBIR, at the lowest level, images are extracted without using semantic information describing their contents. In this case, low-level features are used such as color, texture and shape and some low-level descriptors are applied. In SIIR, at the highest level, image retrieval is based on techniques that allow representing an image with a richer description than low-level descriptors.

Bag-of-visual-words model has first been introduced by (Sivic and Zisserman, 2003) in the case of image and video retrieval. Usually, representing images by vectors of visual words is based on analogies between text and image. Consequently, many effective methods and algorithms inspired from text IR have been applied to the vector of visual words in order to achieve a better retrieval performance.

Nevertheless, it was shown that visual words are not semantically meaningful because in clustering step, these are gathered using only their appearance similarity. So, two visually similar images are not necessarily semantically similar. In order to address these problems, several image retrieval approaches based on ontologies have been proposed (Kurtz and Rubin, 2014), (Allani et al., 2014). The goal is to make images semantic content using annotation terms that are attached to images. However, annotation terms which are used to build ontology, do not guarantee a whole representation or description of images. In this paper, our motivation is to integrate visual vocabulary and ontologies in image retrieval process in order to improve image retrieval accuracy. To perform that, our system focuses on building visual vocabulary and ontologies. During image retrieval process, visual and semantic similarities are integrated.

The remainder of this paper is organized as follows. In section 2, we propose a review of classical image retrieval approaches as well as our motivations. In section 3, we detail our proposal. Section 4 presents our case study. Finally, discussion as well as future works are presented in conclusion.

2 RELATED WORKS AND MOTIVATIONS

In CBIR, bag-of-visual-words model has been widely used for image retrieval, visual recognition, and image classification. Several works using this model have been proposed to provide an efficient visual words in order to apply different image and video processing tasks (Jurie and Triggs, 2005).

In this context, (Sivic and Zisserman, 2003) have presented an approach of object retrieval based on methods inspired from text retrieval. The goal of this work is to retrieve key frames containing a particular object with the ease, speed and accuracy with which Google retrieves text documents containing particular words.

In (Martinet, 2014), a study about visual vocab-

Filali, J., Zghal, H. and Martinet, J.
Towards Visual Vocabulary and Ontology-based Image Retrieval System.
DOI: 10.5220/0005832805600565
In *Proceedings of the 8th International Conference on Agents and Artificial Intelligence (ICAART 2016) - Volume 2*, pages 560-565
ISBN: 978-989-758-172-4

ularies compared to text vocabularies has been proposed in order to clarify conditions for applying text techniques to visual words. To present this study, the author described four methods for building a visual vocabulary from two images collections (Caltech-101 and Pascal) based on two low-level descriptors (SIFT and SURF) combined with two clustering algorithms :K-means and SOM (Self-Organizing Maps). The experiments showed that visual words distributions highly depend the clustering method (Martinet, 2014).

In addition, ontologies based images retrieval approaches have been proposed in order to extract visual information guided by its semantic content (Hyvönen et al., 2003) (Sarwara et al., 2013).

In (Kurtz and Rubin, 2014), a novel approach based on semantic proximity of image content using relationships has been proposed. This method is composed of two steps: 1) annotation of query image by semantic terms that are extracted from ontology and construction of a term vector modeling this image, and 2) comparison of this query image to the others that are previously annotated using a computed distance between term vectors that are coupled with an ontological measure.

In the context of image retrieval based on visual words, when low-level features are extracted, resulting visual words are gathered using only their appearance similarity in the clustering step. Consequently, similar visual words do not guarantee semantic similar meaning. That tends to reduce the retrieval effectiveness with respect to the user. Moreover, in interest points detection step, many detectors can lose some interest points and increase the vector quantization noise. This can result in poor visual vocabulary that decrease the search performance.

Our motivations are to build visual vocabulary and ontologies based on images annotations in order to enhance image retrieval accuracy. The goal is to introduce an image retrieval system which aims to integrate two image aspects: visual features and semantic contents based on images annotations.

Our idea is to combine, during the image retrieval process, similarity between visual words to semantic similarity.

Moreover, the evaluation of our proposal is to achieve two image retrieval strategies:

- A visual retrieval strategy based on visual similarity between visuals words ;
- A strategy based on integrating both visual and semantic similarities. In this case, semantic similarity is based on concepts that are provided from ontologies.

3 VISUAL VOCABULARY AND ONTOLOGIES-BASED IMAGE RETRIEVAL SYSTEM

In this section, we define our visual vocabulary and ontologies-based image retrieval system architecture. Our idea is to build visual vocabulary using low-level features and building ontologies based on concepts that are extracted from images annotations.

As depicted in Figure 1, our image retrieval system is composed of two main phases (online phase and offline phase). The offline phase, which corresponds to the visual vocabulary and ontologies' building phase, is composed of two steps: (1) building the visual vocabulary and (2) building ontology. The online phase, which corresponds to the image retrieval phase, is composed of two steps: (1) query image processing and (2) image retrieval.

In the next section, the different steps of our image retrieval system will be detailed.

3.1 Offline Phase: Visual Vocabulary and Ontologies' Building

Our main idea is to develop an image retrieval system based on building the visual vocabulary and ontologies.

3.1.1 Building the Visual Vocabulary

This step allows to generate visual vocabulary according to three steps: interest points detection, computing descriptors and the clustering phase.

Interest Points Detection: In computer vision many detectors of interest points are developed. In order to produce effective vocabulary we have used the SIFT detector to extract the local interest points because - using this descriptor- a large number of interest points can be extracted from images.

Computing Descriptors (or Feature Extraction): This step consists in extracting features by computing SIFT descriptor for each point which is detected in the previous step.

Clustering: This step consists in clustering local descriptors which are computed in the previous step, the goal is to represent each feature by the centroid of the cluster it belongs. In our case, we have used the K-means algorithm that is the most widely used clustering algorithm for visual vocabulary generation.

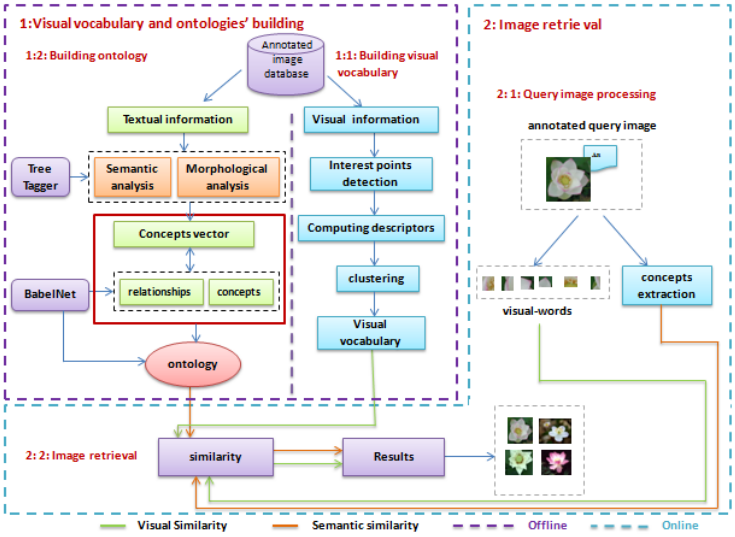

Figure 1: Image retrieval system : main phases and steps.

3.1.2 Building Ontology based on Annotation Files

This process consists in extracting concepts and relationships from annotation files in order to build the ontology. To achieve this goal, a preprocessing step is firstly need. Sub-steps are carried out by :1) extracting image textual information from their annotation files; 2) performing a morphological and semantic analysis in order to get the word's lemmatized form.

- Morphological analysis: it consists in recognizing of the various forms of words using a lexicon (dictionary, thesaurus). The lemmatisation allows the transformation of a word to its canonical form or lemma. In our case the lemmatisation of the annotation files content is ensured by TreeTagger[1].
- Semantic analysis: After the lemmatisation, a filtering step is done. It consists in eliminating empty words.

Then, we need to filter the lemmatized words: only word is a noun, its lemmatized form is added to a vector, this lemma considered as a concept. So, we can detect all existing concepts appearing in annotation files. Finally, we obtain the results concepts that are the input of ontology building process

A lexical resource (BabelNet[2]) is integrated to this step in order to extract concepts and semantic relationships and to enrich our ontlogy. The BabelNet that is organized in synsets, returns all the different meanings attached to words in English language.

Let θ be the ontology which we will build, C_d denote the original concepts which are extracted from annotation, C_{lr} denote the concepts of the lexical resource that is used for extracting relationships, Rt and Rn define taxonomic and non-taxonomic relationships between concepts in C_{lr}. Also we denote SucN and PreN the predecessor and successor concept of a current concept in the hierarchical graph of the lexical resource.

Let consider the sets:

$$C_d = \{C_{d_1},...,C_{d_i},C_{d_{i+1}},...,C_{d_n}\}, i = 1...n \quad (1)$$

$$C_{lr} = \{C_{lr_1},...,C_{lr_j},C_{lr_{j+1}},...,C_{lr_m}\}, j = 1...m \quad (2)$$

$$Rt(X,Y) = \{Rt_1(X,Y),...,Rt_k(X,Y)\}, X \neq Y \quad (3)$$

$$Rn(X,Y) = \{Rn_1(X,Y),...,Rn_l(X,Y)\}, X \neq Y \quad (4)$$

$$X = \{C_d, C_{lr}\}, Y = \{C_d, C_{lr}\} \quad (5)$$

The ontology building process is performed according to the steps:

1) Initialize(θ): add all concepts of C_d in (θ);
2) For each C_{d_i} in C_d and each C_{lr_j} in C_{lr}
   Find $Rt(C_{d_i},C_{lr_j})$;
   Find $Rn(C_{d_i},C_{lr_j})$;
3) If(C_{rl_j} in C_d) then update(θ):
   Create $Rt(C_{d_i},C_{lr_j})$;
   Create $Rn(C_{d_i},C_{lr_j})$;
4) If(C_{rl_j} not in C_d) then update(θ):
   Add C_{rl_j} in(θ);
   Create $Rt(C_{d_i}$,SucN $(C_{lr_j}))$;
   Create $Rt(C_{d_i}$,PreN$(C_{lr_j}))$;

[1]http://www.cis.uni-muenchen.de/∼schmid/tools/TreeTa gger/

[2]http://babelnet.org/

Create $Rn(C_{d_i}$,SucN $(C_{lr_j}))$;
Create $Rn(C_{d_i}$,PreN$(C_{lr_j}))$.

Steps 3) and 4) are repeated until all concepts in C_d are treated and all relationships between them are created.

3.2 Online Phase: Image Retrieval

The retrieval process is based on two steps: query image processing and image retrieval.

3.2.1 Query Image Processing

The aim is to extract the visual information (resp. concepts) from the query image (resp. annotation file). So the query image is represented by a set of visual words or a set of concepts.

3.2.2 Image Retrieval

This step depends on which retrieval strategy that is applied. The retrieval image process can be carried out according to the following strategies:
Visual Retrieval Strategy: In this case, the retrieval is based on the visual similarity between visual words of vocabulary and those that are extracted from the query image. The similarity measures that are used in our case will be defined in the next section.
Strategy based on Combining Visual and Semantic Similarities: This strategy is performed by combining both the visual and the semantic similarity in order to improve relevance of retrieval results. The semantic similarity measures that are used will be presented in the next section.

3.3 Similarity Measures

In order to implement our image retrieval strategies, the similarity measures are computed using visual and semantic similarities.

3.3.1 Visual Similarity

In CBIR field, some works used visual similarity measures (Deselaers and Ferrari, 2011), (Cho et al., 2011). Popular distance measures are used as metric distances like euclidean distance, mahalanobis distance and cosine distance (Zhang and Lu, 2003) and (Cho et al., 2011). In our context we used the euclidean distance to compute similarity between visual words.

$$VisualSim = d(q,r_i) = \sqrt{\sum_{j=1}^{n}(f_j(q)) - f_j(r_i))^2} \quad (6)$$

where

- q : is the vector of query visual words;
- r_i : is a reference visual word i from visual words database;
- f_j : is the jth feature;
- n : is the size of visual vocabulary.

3.3.2 Semantic Similarity

In ontology-based image retrieval, many semantic similarity measures can be used. In this context, many studies used semantic similarity measures in order to increase the performance of semantic retrieval (Hliaoutakis et al., 2006). In our context, semantic similarity between concepts is computed according to the following formula (Patwardhan and Pedersen, 2006):

$$SemanticSim(C_j,C_k) = \eta(C_j,C_k) = \frac{w\vec{C}_j w\vec{C}_k}{||w\vec{C}_j||.||w\vec{C}_k||} \quad (7)$$

where

- C : is the set of concepts related to the query image.
- wC_k : is the concept C_k vector defined in the words space.

4 CASE STUDY

Our main contribution concerns the definition of a retrieval system based on visual vocabulary and ontologies. Our case study is based on ImageCLEF 2008 data-set collection[3] characterized by its diversity. This collection includes:

- 20000 pictures;
- Each image is associated with an annotation file that describes its content.

Figure 2 illustrates the different steps with a specific example related to the given query image. Let's consider a query image composed of the "man", three "women", two "tables" and the "train". So, this image represents different objects. Also, the query image is described by its annotation file.

As depicted in Figure 2, the retrieval process is based on two steps mainly image search. In the image search that depends on the visual vocabulary, when a query image is submitted, a visual vocabulary is generated (Figure 2 Step (7)). After that, visual similarity measure is computed between each visual words of request image and those that are built from image

[3]http://www.imageclef.org/ImageCLEF2008

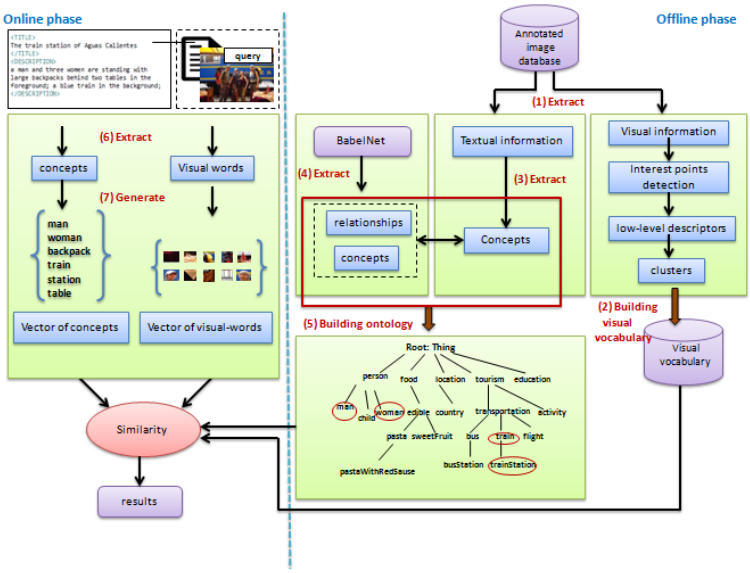

Figure 2: Case study: illustrated process for a given image query.

dataset, according to their similarity, the top ranked images are outputs of the retrieval. In the strategy that based on integrating ontology, visual retrieval is combined to the semantic features based on concepts. During the offline phase, visual and semantic image features are built. This is done by two processes: building the visual vocabulary (Figure 2 Step (2)) and building ontology (Figure 2 Step (5)).

In order to extract concepts the ontology is built based on the annotation files.

After that, the concepts set associated to all of the image dataset is stored. In order to build our ontology, first, ontology must be initialized by adding concepts to it. Next, using relations and hierarchy provided by BabelNet, all taxonomic and non taxonomic relationships related to each concept are extracted (Figure 2 Step (4)). Finally, relationships are added to our initial ontology and related concepts which are do not exist in the initial concepts set, must be added in order to enrich our ontology. For example, using BabelNet, the synsets set that are related to concept ”station” is returned, the results of this example are shown in Figure 3. According to this example, concept ”train” is found, so a semantic relationship between the two concepts ”train” and ”station” is created and added. As depicted in Figure 2, the two concepts are shown in our ontology.

During the online phase, when our query image is submitted to our system, concepts set composed of ”station”, ”train”, ”woman”, ”man”, ”backpack” and ”table” is extracted from its annotation file (Figure 2 Step (6)). Moreover, a set of visual words are extracted that describe low-level representation of this query image. After that, visual words similarity is computed, also, concepts similarity is computed between concepts that are extracted from the query image and those which form our ontology. As depicted in Figure 2, many concepts which represent semantic content of this query image like ”station”, ”women”, ”man” and the ”train”, are shown clearly in our ontology. In this case, three image retrieval strategies can be done: 1) this strategy consists in first, doing visual search based on low-level features; second, keeping the tops relevant images; then, applying the semantic retrieval using ontology, 2) the second strategy begins with the semantic retrieval, only images results are then used for performing visual search , and 3) the third strategy consists to do both visual and semantic retrieval separately and the intersection of each set of top retrieved images provides the research result of our system.

On the contrary of classic approach based on visual vocabulary, integrating ontology and combining visual retrieval to semantic features could improve research performance by getting more coherent results for given image request. Our novel system combines two features aspects that are achieved by building visual vocabulary and ontologies. Integrating ontologies ensure high level semantic processing that influences the research quality.

5 CONCLUSIONS

In this paper, we introduced our image retrieval system based on building visual vocabulary and ontologies.Our proposed system is focused on two image as-

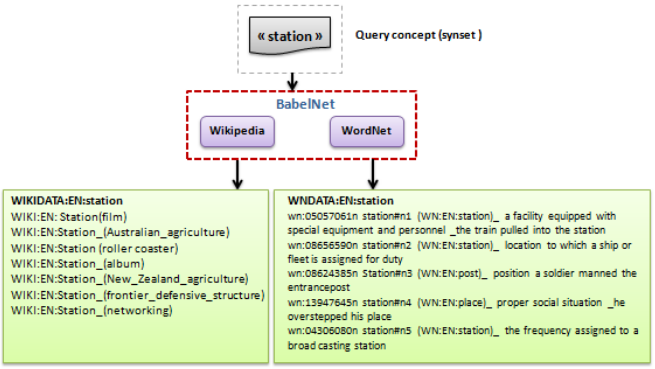

Figure 3: Example: BabelNet results.

pects: visual and semantic features. During the offline phase, visual vocabulary is built in order to describe image database by their own visual content. Moreover, this phase can be performed by extracting concepts and relationships between them from annotation files in order to build ontologies that are then used in retrieval process. The ontologies are enriched by the concepts and relationships that are extracted from the BabelNet's lexical resource. The case study shows the feasibility of our proposal.

In future works, we will evaluate our proposed system. In order to achieve this goal, we will compare retrieval results from our image retrieval strategies based on combining both visual and semantic similarities with the classical content based image retrieval based on visual vocabulary.

REFERENCES

Allani, O., Mellouli, N., Baazaoui-Zghal, H., Akdag, H., and Ben-Ghzala, H. (2014). A pattern-based system for image retrieval. In *The International Conference on Knowledge Engineering and Ontology Development.*

Cho, H., Hadjiiski, L., Sahiner, B., and Helvie, M. (2011). Similarity evaluation in a content-based image retrieval (cbir) cadx system for characterization of breast masses on ultrasound images. *Medical Physics, The International Journal of Medical Physics Research and Practice*.

Deselaers, T. and Ferrari, V. (2011). Visual and semantic similarity in imagenet. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1777–1784. IEEE.

Hliaoutakis, A., Varelas, G., Voutsakis, E., E.Petrakis, and E.Milios (2006). Information retrieval by semantic similarity. *International Journal on Semantic Web and Information Systems*, 2:55–73.

Hyvönen, E., Saarela, S., Samppa, Styrman, A., and K.Viljanen (2003). Ontology-based image retrieval. In *WWW (Posters)*.

Jurie, F. and Triggs, B. (2005). Creating efficient codebooks for visual recognition. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 604–610. IEEE.

Kurtz, C. and Rubin, D. (2014). Using ontological relationships for comparing images describing by semantic annotations. In *EGC 2014, vol. RNTI-E-26, pp.609-614*, pages 609–614.

Martinet, J. (2014). From text vocabularies to visual vocabularies: what basis? In *9e International Conference on Computer Vision Theory and Applications (VISAPP),pp. 668-675, January 2014, Lisbon, Portugal.*

Patwardhan, S. and Pedersen, T. (2006). Using wordnet-based context vectors to estimate the semantic relatedness of concepts. In *EACL 2006 Workshop on Making Sense of Sense: Bringing Computational Linguistics and Psycholinguistics Together, pages 18.*

Sarwara, S., Qayyuma, Z., and Majeedb, S. (2013). Ontology based image retrieval framework using qualitative semanticimage descriptions. In *17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems -KES2013*, page 285 294. Procedia Computer Science 22.

Sivic, J. and Zisserman, A. (2003). Video google: A text retrieval approach to object matching in videos. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on.*, pages 1470–1477. IEEE.

Zhang, D. and Lu, G. (2003). Evaluation of similarity measurement for image retrieval. In *Neural Networks and Signal Processing, Proceedings of the 2003 International Conference on (Vol. 2, pp. 928-931).* IEEE.