

Ontology Driven Content Based Image Retrieval

Adrian Popescu
CEA/LIST - LIC2M

18 route du Panorama
92260 Fontenay aux Roses, France
+33146548013

adrian.popescu@cea.fr

Christophe Millet
CEA/LIST - LIC2M

18 route du Panorama
92260 Fontenay aux Roses, France
+33146548137

milletc@zoe.cea.fr

Pierre – Alain Moëllic
CEA/LIST - LIC2M

18 route du Panorama
92260 Fontenay aux Roses, France
+33146549619

pierre-alain.moellic@cea.fr

ABSTRACT

Content based image retrieval (CBIR) methods are proposed as alternative or complementary solutions to keyword-based picture search. However these techniques mostly rely on low-level descriptors similarity between different items and when one uses such an application to find pictures, the proposed answers are often not conceptually similar to the query. In this paper, we describe Retrievo, an image retrieval (IR) system that allies CBIR techniques and semantics in order to better fit the users' expectations when querying an image database. The dataset is structured employing a term hierarchy, which is used to control the conceptual neighbourhood where similar items are searched. Only the leaf terms of the hierarchy have associated image sets but, with the use of the type-subtype relation between nodes, pictures are indirectly associated to all the concepts in the hierarchy and the system can propose localized IR processes, which associate low-level and conceptual similarities (on different levels of generality). We model a real-world situation by using pictures gathered from the Internet. The ontologically controlled IR method proposed in this paper is compared to classical CBIR functioning and we show that the introduction of a hierarchical structure improves precision results for the system.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *retrieval models, search process, information filtering.*

General Terms

Algorithms, Experimentation

Keywords

Content based image retrieval, semantics, ontology, WordNet

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIVR'07, July 9-11, 2007, Amsterdam, The Netherlands.

Copyright 2007 ACM 978-1-59593-733-9/07/0007 ...\$5.00.

1. INTRODUCTION

The considerable amount of still images available on the Internet, in public or private picture databases require the development of efficient retrieval techniques. There exist two main IR paradigms: keyword and content based. The keyword based search is computationally efficient (it simply treats preindexed character chains) and is exploited in very large scale applications like Internet picture search engines (Google, Yahoo!, Picsearch, Ask etc.). The number on indexed images on the Web exceeds 1 billion and it would be too expensive to manually associate keywords to such a volume of data. The image search applications are based on automatic annotation algorithms which exploit the text around the picture. Not all spotted keywords are related to the visual content [16] and when searching for responses to illustrate a query, the obtained results are often unsatisfactory. Here, the retrieval process is text based and there is no content analysis performed.

On the contrary, classical CBIR systems like those described in [8], [11], or [15] rely on picture analysis and do not employ high-level semantics related to it. The only image parameters which can be automatically computed and compared are low-level ones like colour, texture and/or shape. The obtained similarities between items make sense from a machine point of view but are often meaningless for a human. Picture analysis is a computationally complex task and it is currently impossible to propose content based similarities processing in very large scale datasets in real-time conditions. Anecdotaly, the largest available CBIR demo, Cortina [8] runs on a dataset of about 11 million items. The lack of human understandable semantics of results and scalability problems are two important difficulties on CBIR systems' road to large scale adoption.

An approach which combines minimal linguistic information about pictures and content based similarities is discussed in this paper. We extract the term hierarchy ranging under *placental* in WordNet [7] and separate the leaf synsets in this ontology. The hierarchy contains a total of 1113 synsets (synonym sets), with 841 leaves.

The latter are used to query the Web for pictures related to terms in these synsets and, once the dataset is constituted and filtered, to associate them to the ontology. The Internet is a rich and interesting to exploit resource but the results retrieved when searching visual representations related to a concept are sometimes disappointing. With the use of specialized terms in WordNet, the retrieval precision is improved compared to the

situation when familiar category names are used [10]. Nonetheless, it is important to filter unwanted items in order to obtain as much accuracy as possible in the resulting database. We observed that noisy answers are often cliparts or contain human faces, and employ a filter bank which designed to eliminate them.

The hierarchy has picture sets associated to leaf synsets and, via the type – subtype relation, these sets are indirectly related to higher concepts in the hierarchy. With the use of the ontology, a structured database is obtained and it is possible to perform localized content based retrieval. The system proposes image search at different levels of generality, determined by the position of the root term of the current subhierarchy. For example, if one asks for pictures similar to a particular representation of *eastern grey squirrel*, the application proposes results in term hierarchies defined by the concept itself or its parents: *tree squirrel*, *squirrel*, *rodent* and *placental*. It is to be noted that, when using placental as root term, the retrieval process is purely content based.

The framework discussed here addresses three key problems in CBIR:

-the lack of human understandable semantics – with the use of structured linguistic information, when querying the system it is possible to control the conceptual neighbourhood where similar pictures are searched.

-scalability – associating images to leaf terms is significantly more economic than to propose a representation for all the concepts in the hierarchy, while not loosing coverage. Equally important, for very large scale applications, it is possible to restrict the retrieval process to real-time manageable conceptual neighbourhoods.

-interactivity – one frequent critic related to CBIR systems is that they are not self-explanatory and offer very little interaction means to the users. An organized database enhances the interaction possibilities, while providing easily understandable information about the states of the system.

The remainder of this paper is organized as follows: Section 2 reviews relevant related work. Sections 3 and 4 respectively present the conceptual side of RetrieveOnto and the image database constitution process. In Section 5, we provide an overview of the obtained system. Before concluding, we present an evaluation of our method against classical content based retrieval.

2. RELATED WORK

A detailed survey of CBIR techniques which include semantics (SCBIR) can be found in [6]. The authors propose a classification of SCBIR methods in five main categories:

1. employing ontologies to define high-level concepts
2. using machine learning techniques to associate low-level features with query concepts
3. introduction of relevance feedback into the system in order to account for the users' actions
4. generating semantic templates to assist high-level IR
5. using both visual content and surrounding text from the Web to assist IR

Our approach is closest to the fifth category, as it jointly uses visual similarity and textual annotations related to Web pictures but, to our knowledge, there exists no system of this kind using a dataset which is directly structured by an ontology.

Cai et al. [2] present a complex framework for IR. They analyze both the visual representation and the surrounding text and produce a three-folded representation of a Web picture including: visual, textual and link information. They propose a two-step clustering method: first, a text and link based clustering is used to reduce the search space to semantically related images sets and, second, a picture clustering algorithm is employed to group visually resembling images. During the textual clustering phase, a combination between a cosine measure and spectral clustering is proposed to obtain affinity matrix. The authors of [2] underline an important shortcoming of their method: the reduced amount of text associated to a part of Web images. They propose the use of link information to complete the high-level clustering phase in their system. In RetrieveOnto, we propose a simpler framework for textually grouping pictures: specialized concepts from a hierarchy are used and semantic grouping of images is realized using the inheritance relation between concepts in the hierarchy.

Ferecatu et al. [3] introduce an IR framework which combines the use of ontological information (extracted from WordNet) and of low-level descriptors of a visual representation. Two corresponding feature vectors are created and evaluated to propose similar items. WordNet is used to extract inheritance paths relating keywords to some key-concepts in the hierarchy. A key component of the application is the relevance feedback mechanism employed in a machine learning process. The system in [3] works on a manually annotated database of less than 4000 items and, with the introduction of a conceptual based feature vector related to a picture, the obtained results significantly improve when compared to a pure content based retrieval procedure. The IR application proposed in [3] is quite complex and its extension to large scale dataset would be a complex task. A human intervention is required to define key concepts which characterize the images and to guide the learning process associated to classes.

In [16], Yavlinsky et. al present a system (Behold) which combines classical keyword search and a visual keyword vocabulary. This structure contains a limited number of terms (57) which are automatically associated to image content. The picture repository in [16] contains over 1 million images and significant improvement of results is reported for the images annotated with terms from the visual vocabulary. As machine learning techniques are used to define the visual lexicon, an important difficulty related to the approach in [16] is the extension of the vocabulary. The role of keywords in Behold is significantly different from the one they have in our framework as there is no semantic structure which organizes them.

The use of ontologies to define high-level concept properties is proposed to ameliorate the results of CBIR systems [6]. These ontologies associate low and high-level information about image components (i.e. colour), trying to abstract object characteristics in a bottom-up fashion and to use them in the retrieval phase. An interesting tentative to jointly use ontologies and CBIR techniques is reported in [9]. In this paper, the authors use a domain ontology to enhance picture region recognition. The ontology contains inter-conceptual and spatial information relative to the included objects. This type of approach is related to ours in that it uses ontologies in IR, but the role of these knowledge frameworks is different. The mentioned work proposes to reduce the semantic gap via a direct association of high and low-level parameters of images and try to model their content

using knowledge frameworks. We do not analyze the visual content in order to attach textual labels and propose the utilization of a concept hierarchy to structure the repository.

We briefly present hereafter some classical CBIR systems. Cortina [8] is an application which exploits low-level characteristics (colour, texture and edge) of Web images to propose a similarity search in an 11 million items dataset. The initial query is keyword based and there is no high-level semantics involved in the retrieval process. WebSeek [11] is an application somewhat similar to Cortina which uses an image and video database including more than 600000 items. A noteworthy difference between the two systems is that, in WebSeek, an ad-hoc hierarchy is proposed as initial browsing option. The retrieval process in WebSeek is uniquely based on low-level visual parameters.

Simplicity [15] is another well-known CBIR system which employs a proprietary dataset composed of around 60000 objects. Several querying options are proposed to the user: random proposition of items; a selection of items displayed by the system, a drawing interface and the comparison to a depiction proposed by the user. Once again, the retrieval process is purely low-level.

Tiltomo [13] is a recently developed application which uses pictures spidered from an online collaborative application (Flickr). These items are manually annotated by the users of Flickr and Tiltomo proposes two search options: textual thematic and low-level visual based similarities. The CBIR facility is uniquely based on texture and colour analysis and the tags related to the pictures are not exploited.

For all classical CBIR systems, the obtained results lack conceptual coherence as the images are grouped using only low-level visual descriptors. This is an important drawback towards the adoption of such systems by a large public. We discuss in the following some possible ways to improve CBIR results using high-level semantics and to provide more understandability.

3. THE CONCEPT HIERARCHY

The picture database employed in Retrievo is structured using a term hierarchy automatically extracted from WordNet [7]. We described elsewhere [10] a translation of the WordNet lexical database into a Semantic Web compliant language, OWL, and use a simplified version of this transformation in the application described in this paper. For demonstration purposes, we retained a subhierarchy containing the terms ranging under *placental* in WordNet. It is to be noted that the inheritance system in WordNet is not a scientific one and is close to commonsense knowledge (e.g. *dog* has *puppy* as immediate subconcept). Another characteristic of the hierarchy is that some parts are far better developed than others. There are 144 leaf nodes under *dog* and only 10 subconcepts of *dolphin*. The subconcepts of the two categories are not exhaustively described in WordNet but this example reflects the fact that there is more commonsense knowledge associated to dogs than to dolphins. The main rationale for choosing a commonsense hierarchy over a scientifically valid one is that the proposed application is aimed to be a non-specialized one. The depth of *placental* hierarchy ranges from 1 to 8. For example, *livestock* is a terminal node which inherits directly from the root, while *Brown Swiss* is a leaf subconcept of *placental* passing through: *dairy cattle*, *cattle*, *bovine*, *bovid*, *ruminant*, *even-toed ungulate*, and *ungulate*.

The conceptual structure includes a total of 1113 nodes, with 841 leaf terms. The latter are different from other nodes in that they have associated picture sets. For polysemic concepts, like *Angora*, each meaning is represented as a separate class in the ontology. This property is particularly interesting from an IR point of view because each sense of a term points towards a distinct category in the world.

The role of the term hierarchy is to control, in a humanly understandable fashion, the region of the database where similar items are retrieved. For example, one can ask for items which are close to a representation of *Brown Swiss* in conceptual neighbourhoods determined by this class or by any of its parents in the hierarchy, up to *placental*. With the increase of the generality of the root term of the subhierarchy, the size of the image set where similar representations are retrieved gets bigger as well as the semantic distance between the included leaf nodes.

4. THE PICTURE DATABASE

CBIR applications use two main types of databases [6]: standardized (i.e. Corel, Columbia) or formed by querying the Web. The former are preferred in a majority of CBIR applications [6] and are used in systems like SIMPLICITY [15], IKONA [1], PIRIA [4], while the latter are employed in Cortina [8], WebSeek [11] or Tiltomo [13]. Standardized databases contain good quality pictures (i.e. the depicted objects are well represented). The obtained results are easy to evaluate because there is no noise associated to the annotation (the association image content – keyword is reliable) but these databases do not always reflect real-world conditions. This is especially true when one wants to develop Internet related applications, where pictorial representations are of variable quality and the keywords associated to them are not always related to the content. Although noisy and rendering poorer results, the repositories constituted using Web picture give a better idea about the capabilities of the system when dealing with pictures from heterogeneous sources. We decided to use Internet to constitute a test dataset.

4.1 Image Gathering

We employ an existing search engine, Ask, to populate the database. The choice of this particular application was determined by its higher precision results when compared to other similar applications like Google, Yahoo! or Picsearch. A comparative test was ran on a series of 20 concepts (with 50 images per query) and, for Ask, the rate of correct image content – keyword association was around 80%. For the second best system, Picsearch, the same parameter scarcely exceeded 70%.

Leaf terms in the concept hierarchy were used to form a list of queries which were submitted to recuperate pictures from Ask. In WordNet, synonyms are grouped in synsets and, when several terms point towards the same entity in the world, queries are launched using each member of a synset. If a query term is polysemic, a disambiguation procedure is followed in order to distinguish between the different depictions of a same linguistic concept. For example, the *German shepherd* is also called *German shepherd dog*, *German police dog*, and *Alsatian*. The first three names of the concept are monosemous, while *Alsatian* is polysemous. The disambiguation consists in formulating a composed query, using the immediate parent of a polysemous term. In this case, we obtain *Alsatian shepherd dog*.

The Internet is a rich resource but it does not contain pictorial representations for all the concepts in the language. Moreover the number of pictures associated to different concepts is highly variable. We initially collected over 33000 images. After the elimination of invalid links and invalid files, there are 31287 items left. The use of image filtering techniques (see Section 5) reduces the database to 25470 items, distributed over the leaf nodes of the hierarchy. The mean number of pictures in a class is 30.3, with a standard deviation of 23.8. The minimum number of items associated to a terminal node varies between 0 and 147. The terms which do not have a pictorial representation are either rare ones, like *Pteropus capestratus* or *baronduki*, or secondary senses of known terms, like *doe* or *yearling*. For the latter, the primary meanings do not belong to the placental hierarchy. Well represented nodes are related to familiar concepts like *hippopotamus* or *grizzly*.

5. IMAGE PROCESSING

We are interested in retrieving depictions of animals. Therefore, before the indexing phase, we remove any item that is not related to our query, namely those containing faces and representations of computer drawn cliparts or scanned texts.

5.1 Removing faces

Photographs of faces are a very popular subject among digitized pictures. The consequence is that any keyword used in a web search engine is likely to return representations of faces (see Figure 1). These images can be related to the query, as for example if we are looking for representations of a *Siamese cat*, and find a picture of someone holding his pet. They can also be indirectly related, such as photo of a *whale* specialist, or photos of someone whose nickname is the name of what we queried, and it has been associated to the image.

All these representations are considered as noise, since we would like to show pictures depicting only animals. In order to eliminate items containing faces, we need a face detector. Then, images where we detect a face are removed from the database. The face detector we used is the multi-stage AdaBoost detector proposed by Viola and Jones in [14] with the improvements of [5].



Figure 1. Examples of retrieved images with faces for queries *aardvark* (left) and *peba* (right).

If the total area of the detected faces in a given picture is small (we empirically defined smallness as being less than 5% of the surface), the result of the detection is considered a false positive, and that item is not excluded from the database.

5.2 Removing cliparts and scanned texts

Other item types we want to remove are cliparts and scanned texts. Cliparts are sometimes related to the query, but we are interested only in browsing for photographs. We believe that cliparts and photographs are two different categories of pictorial

representations that should be considered separately, because they require different processing algorithms.



Figure 3. Examples of cliparts obtained with the queries “golden mole” (left) and “livestock” (right).

Scanned texts are another kind of items that arise when querying the Web, mostly because some animals represented in WordNet are referred to by their scientific name. Using this as a keyword in a web image search engine retrieves scanned scientific publications.

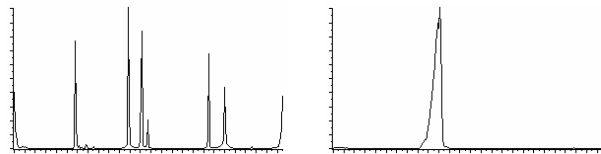


Figure 2. Luminance histograms of a clipart (left) and a photograph (right) containing the same number of colours.

The algorithm we developed to separate photographs from cliparts and scanned texts is based on the shape of histograms. In Figure 2, we notice that, even if a clipart and a photograph have the same number of colours, their luminance histograms have different noticeable properties. The histogram for a clipart is more discrete and made of peaks, whereas the histogram for a photograph is more continuous.

Therefore, the criterion we have chosen is to locate the maximum of the histogram, compute the standard deviation around that maximum, and then use a threshold to decide whether it is a clipart or a photograph. This threshold can be experimentally determined to optimize classification results on a database of sample images. Typically, the peaks of cliparts will have a low standard deviation, whereas the standard deviation for a photograph will be higher.

If we use this algorithm on the whole picture, it classifies photographs with a uniform frame as cliparts. An improvement consists in first dividing the image into 16 equal rectangular parts. The standard deviation is computed on each part, and the threshold for the classification is applied on the higher standard deviation. On a database of 11250 photographs and 5402 cliparts from the Internet, this algorithm correctly classifies 99.8% of the photographs, and 93% of the cliparts. It proved to work better than a trivial algorithm that would just consist of counting the number of colours in the picture.

5.3 Indexing images

We index the dataset with the border/interior pixel classification algorithm proposed in [12]. It first quantizes each R, G and B component into 4 values. Then, pixels are classified into border pixels or interior pixels: a pixel whose 4 neighbours have the same quantized colour is called interior, and border otherwise. Eventually, two 64 bins RGB histograms are built: one for border pixels, and one for interior pixels.

When indexing a picture, we want to limit the influence of the context, ignoring as much as possible any background surrounding the represented object. Since automatic segmentation is a hard task, and does not ensure the separation of the object from its context, we propose simply to only index the centre of the image, based on the hypothesis that the object is usually placed there. We consider only the pixels contained in the central window of the picture, covering a quarter of its surface.

6. THE RETRIEVONTO SYSTEM

When combining the conceptual hierarchy, the processed dataset and a user interface, a functional CBIR system is obtained. The application uses the high-level semantics of an inheritance system to control the retrieval process and to propose browsing variants to the user. A frequent critic [6] related to CBIR applications is that they do not offer enough interaction means to the user and that they are not self-explanatory. RetrievoNto was designed to cope with these problems and provide an easy to use user interface.

The main states of the system are the query mode and the answers page. It is possible to navigate between these states or to change the parameters of one of them in order to obtain different configurations of the system. The two states and the interaction means deployed in RetrievoNto are discussed in the following three sections.

6.1 Query Mode

There are two query modes offered in RetrievoNto:

- a set of random items depicting different leaf concepts in the hierarchy – this is the default presentation mode of RetrievoNto and if the user is invited to click on a picture in the set to get similar ones from the database. If this option is chosen, a response page appears. Alternatively, one can choose to have other items presented or to pass in a conceptual browsing mode.

- conceptual browsing – in this mode, a list of 30 randomly chosen leaf terms is presented to the user and, if one is selected a set of pictures representing it are displayed. The terms presented to the user have different sizes, in direct correlation with the number of associated pictorial representations on the Web. The idea behind this representation is similar to that applied in Flickr¹ and tends to favour pictorially well represented terms over the others. The user can click any of these images and a response page appears.

6.2 Answers Page

Once the query mode is selected, we get a selection of pictures associated to different concepts or to a same leaf term. When clicking an item, the 19 (or less if the number of pictures in the dataset is not sufficient) most similar answers in a defined conceptual neighbourhood appear. By default, the search is performed among the items belonging to the same concept. Options to perform a retrieval process in a larger neighbourhood are offered either by using a button that moves the root concept one step up in the hierarchy or by clicking any of the parents of the leaf term which are displayed just above the image responses.

In Figure 4, the query picture is located on the upper-left corner. Nineteen similar items belonging to the same leaf node are

displayed in an ordered manner. *Procyonid* is the direct hypernym of giant panda. A CBIR process using the same query as in Figure 4, is performed in the subhierarchy defined by *procyonid* and the responses are presented in Figure 5. In Figure 6, we depict the results of a classical CBIR process (i.e. similar items are retrieved in the entire database).

In Figure 5, similar items are associated to the following leaf nodes: *giant panda*, *lesser panda*, *common racoon*, *coati*, and *crab-eating raccoon*. These concepts are all subtypes of *procyonid*.



Figure 4. Answers page for a query (upper-left corner) with a picture of a *giant panda*. All images belong to the same class.



Figure 5. Answers page for a query with a picture of a *giant panda*. The images belong to the subhierarchy defined by *procyonid*.

The images in Figure 6 belong to a variety of species of *placentals*. Depictions of *King Charles spaniel*, *cave myotis*, *mule*, or *mountain gorilla* are in the answers set, but these classes are weakly related to giant panda. The conceptual neighbourhood in Figures 5 and 6 are larger than the one in Figure 4 and the

¹ www.flickr.com/explore

quality of the obtained results decreases. Representations which are not visually similar from a human's point of view appear in the response set in Figure 5 and prevail in Figure 6. A formal evaluation of effect of enlarging the conceptual hierarchy where similar images are retrieved is provided in Section 7.

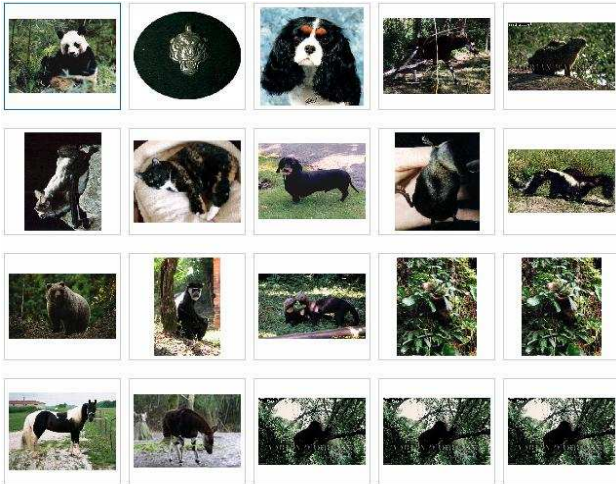


Figure 6. Answers page for a query with a picture of a *giant panda*. The images are selected from the entire database (classical CBIR).

6.3 Browsing the results

The user interaction interface is a problem which is often disregarded when building CBIR systems. The use of a structured picture repository in IR enables augmented browsing possibilities on the user side, while keeping the interface easily understandable.

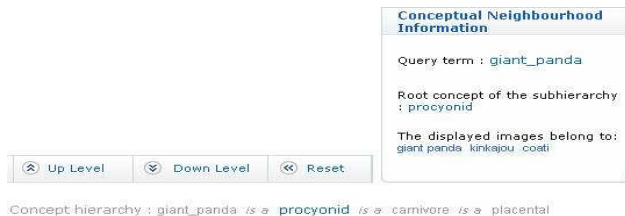


Figure 7. Global view of an answers page

A grouped view of the interaction means in *RetrievOnto* is offered in Figure 7. We note that these means are of two types: buttons and conceptual links. The buttons appear in the results pages: “Up level”, “Down level”, and “Reset”. The first two enable the user to visualize results in larger, respectively narrower conceptual neighbourhoods, while the latter provides a link towards the main page.

The “Down level” and “Up level” buttons are disabled when the retrieval processes are performed in the most specialized, respectively most general neighbourhood. An explanation box concerning the three buttons is displayed on the right side of the response page. The current root concept is highlighted (*procyonid* in Figure 7) and the user is always informed about the generality degree of the conceptual neighbourhood where his search is performed.

The conceptual links are situated in two areas of the answers page: above the results and in an information box on the right of the page. The parents in the concept hierarchy (Figure 7, lower side) are clickable at any moment. The information box (Figure 7, right side) appears on the right side of the response page and it contains information concerning: the leaf term to which the query is associated, the current root of the subhierarchy and the list of concepts that are represented among the presented similar pictures set. It is possible, at any moment, to click on a concept name in the box and to get a selection of visual representations (a state that is similar to the conceptual browsing mode). We mention that all the similar pictures can be clicked and, if so, they are considered as queries and a response page containing similar items appears.

7. EVALUATION

Two measures are classically employed to evaluate the IR applications: precision and recall. The recall is important in small scale databases, where it is important to retrieve as much correct answers as possible. This parameter becomes less important for large scale and heterogeneous datasets, like Internet images, where it is impractical to present all the pictures which are close to a query. Moreover, the database usually contains wrongly annotated images and this fact would fake the recall measure. Precision is an important parameter in both situations because it defines the quality of the presented answers irrespective to the size of the database.

We performed three types of tests which are discussed hereafter. The first one addresses the overall quality of the database. Thirty classes (see a complete list in figure 8) were selected as to cover as much as possible the different branches of the hierarchy (i.e. to represent a large panel of animals represented in the placental ontology) and, for each class, twenty randomly chosen items were presented to the evaluator. For each picture he was asked to decide if it is representative for the given class. The species represented by the leaf nodes are not necessarily familiar to the evaluator and it was necessary to provide some positive examples in order to facilitate his task. 86% of the individuals in the evaluation set were judged representative. We remind the user that the images were collected using the Ask search engine and observe that the results obtained here are coherent with those described above.

A second test was meant to determine the efficacy of the filtering techniques presented in Section 5. The test conditions were similar to those described above and the evaluated dataset contained 200 pictures (both drawings and faces). We found that 35% of the items in this set were not representative for the respective classes, a percentage that is to be compared with the 14% obtained above. The increased number of wrong pictures in the drawings and faces list proves the utility of the applied filtering techniques but further refinement of the filtering bank is needed.

After the evaluation of the global quality of the database, it is necessary to assess our ontology driven approach to IR. The baseline for this comparison is the classical CBIR compartment of the system, obtained when similar items to a queried one are searched in the entire database (that is when the ontology does not play any role). For each image, the twenty most similar answers (from a machine's point of view) are selected using all the

conceptual neighbourhoods defined by the set of parents of a leaf term in the ontology.

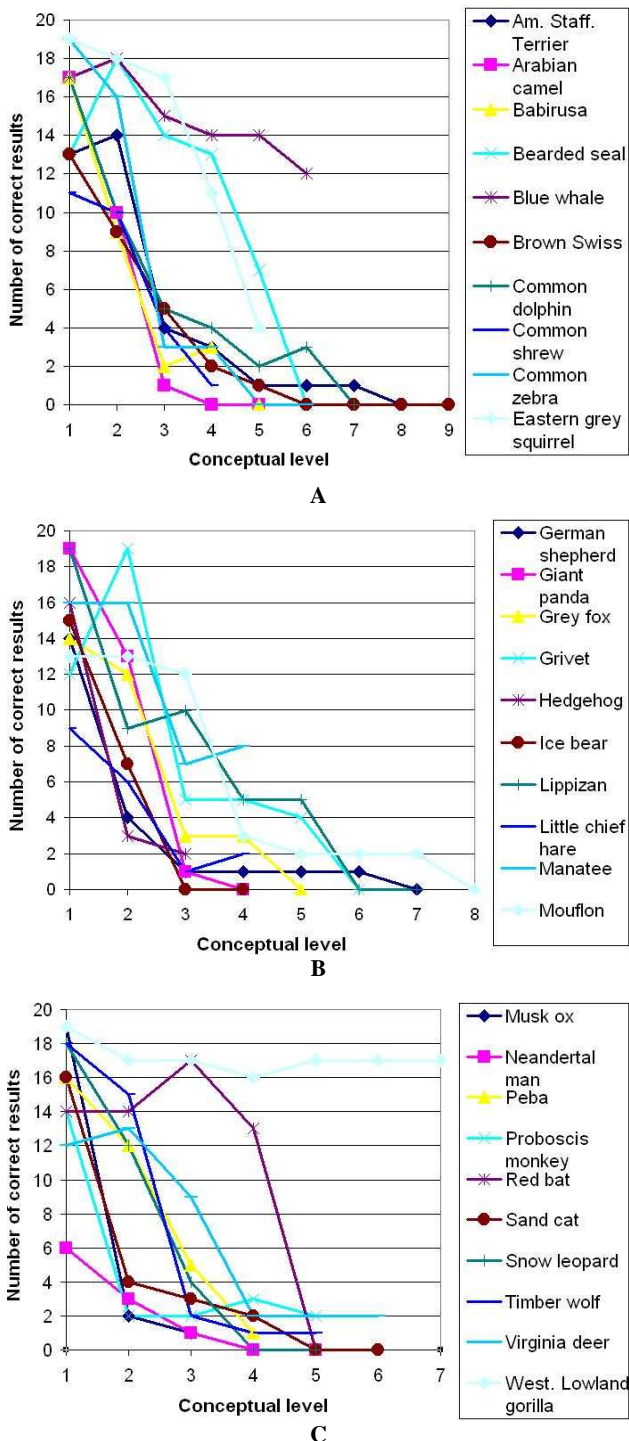


Figure 8. Precision results for 30 queries (fig. 7A – 1 to 10; 7B – 11 – 20; 7C – 21 to 30). Conceptual level 1 stands for a retrieval among the images belonging to a same leaf node. Levels 2 to 9 represent larger conceptual neighbourhoods.

For the thirty concepts, the depth of the hierarchy varies between 3 and 9. For example, a picture representing a *Brown Swiss* cow will be evaluated on 9 levels, the class itself being the most restrictive and the entire repository (when the root concept is *placental*) the largest. In set of 30 all the queries are relevant for the associated keyword.

The results of the evaluation are presented in Figure 8 (A, B, and C). For each query, the number of relevant items (out of 20) is provided for all the levels in the hierarchy. Conceptual level 1 stands for a content based retrieval performed only among items belonging to the same leaf node as the query. The rightmost level of each curve stands for a search in the entire database (the reader is reminded that the depth of the hierarchy is variable and this explains the differences between the number of levels per query).

The results in Figure 8 show that the best precision is generally obtained in narrow conceptual neighbourhoods (left of the figures), while the classical CBIR functioning mode renders the poorest results (right of the figures). With two exceptions *blue whale* (fig. 8A) and *western lowland gorilla* (fig. 8C), where the answers are adequate for all the levels of the ontology, the results obtained when searching the whole database are by far worse than those obtained when using the hierarchy. The curves in figure 8 and the examples in Figures 4, 5, and 6 confirm an intuitive supposition, namely that a decreased role of semantic control of the IR process (retrieval in larger conceptual neighbourhoods), results in poorer quality results.

It is noteworthy that the database contains images which are not representative for the leaf nodes to which they are associated. The noisy items in the dataset limit the precision of the search at all levels in the hierarchy. The results for the narrowest conceptual neighbourhood (level 1) presented in Figure 8 correlate with those we reported above, when testing the precision in the picture repository. The mean precision for the conceptual level 1 is 76%, a depreciation of 8% when compared to the mean precision in the database. The difference is explained by the fact that the visual similarity test (Figure 8) accounts for both the mean precision in the database and the distance between the contents of the compared pictures.

It is also remarkable that the results often floor when the conceptual neighbourhood becomes larger than our common knowledge about a species of animals. For example, the *German shepherds* (fig. 8B) are easily separable from other *shepherd dogs* (the immediate parent of the node). It is more difficult to separate the *American Staffordshire terriers* (fig. 8A) from other *bulldog terriers* (the immediate parent of the leaf), but it is easy to separate these last from other *terriers* (parent class of bulldog terrier). For *German shepherd*, the results floor between this class and its immediate hypernym, *shepherd dog*, while for *American Staffordshire bulldog terrier*, the results become worse when passing from *bulldog terrier* to *terrier*. Similar effects appear, among others, for classes like *common zebra* (fig. 8A - separation line between *zebra* and *equine*), *eastern grey squirrel* (fig. 8A - the results floor between *squirrel* and *rodent*), or *grey fox* (fig. 8B - separation line between *fox* and *canine*). This observation is important if considerably larger image repositories will be dealt with as it is possible to stop the retrieval process at a predetermined intermediary level and not to go up to the root of the hierarchy.

Some critics may apply to the evaluation method. One of them is related to the subjectivity of the picture evaluation. There are

three factors which discard this critic: the evaluator was uninformed about the used IR method; the high number and diversity of evaluated items; the significant differences in performance when using a structured database compared to a classical CBIR functioning of the system

The values in Figure 8 are (yet another) proof for the need for semantics in IR. The introduction of a conceptual hierarchy to structure the picture database and its association with image processing techniques clearly improves the results of the retrieval process, while not being computationally expensive. The main computational burden still lies in the image processing steps. This observation is important when one wants to scale up a CBIR system. In a related work [3], the scaling process is one unsolved key problem. With our approach, it is possible to include larger picture repositories in the system or to use wider ontologies. Currently, the system works in real time (it may take up to a few seconds to answer a question in the worst case) but not much effort has been spent on algorithmic optimization of the application.

8. CONCLUSIONS

In this paper we presented a novel approach to image retrieval, which associates the use of an ontology and of low-level picture descriptors. The originality of the approach comes from the use of the conceptual hierarchy to structure the dataset. We equally discussed the constitution of a picture database using an Internet search engine and leaf nodes from the conceptual hierarchy and the image processing techniques employed to filter and index the picture. An evaluation of our approach against state of the art techniques was performed and it showed that, with the use of minimal semantic information (a term associated to each photo), the results of a CBIR process are fairly improved.

The results presented in this paper encourage us to develop our approach. In the future, we will focus on the extension of the image database, correlated with the use of a wider ontology. Another line of work is represented by the amelioration of the image processing techniques, notably the introduction of automatic segmentation and the improvement of the face and clipart filtering techniques. We are equally interested in testing the reaction of users when presented with increased interactivity in CBIR systems which combines semantics and image processing techniques. In a later stage, the work presented here is to be integrated into a more complex picture retrieval tool, combining keyword search and content based retrieval.

9. ACKNOWLEDGEMENTS

We thank Sofiane Souidi for having developed the PHP interface of the application.

10. REFERENCES

- [1] Boujemaa, N., Fauqueur, J., Ferecatu, M., Fleuret, F. Gouet, V., Le Saux, G., and Sahbi, H. IKONA: Interactive Generic and Specific Image Retrieval. In *Proc. of Int. Workshop on Multimedia Content-Based Indexing and Retrieval (MMCBIR'2001)* (Rocquencourt, France).
- [2] Cai, D., He, X., Li, Z., Ma, W.-Y., Wen, J.-R. Hierarchical clustering of WWW image search results using visual, textual and link information. In *Proc. of 12th ACM Int. Conf. on Multimedia* (New York, NY, USA, 2004), 952 – 959.
- [3] Ferecatu, M., Boujemaa, N., and Crucianu, M. Semantic interactive image retrieval combining visual and conceptual content description. To appear in *ACM Multimedia Systems Journal*, 2007.
- [4] Joint, M., Moëllic, P.-A., Hède, P., and Adam, P. PIRIA: A general tool for indexing, search and retrieval of multimedia content. In *Proc. of SPIE Image processing: algorithms and systems* (San Jose, California, January 19 – 21, 2004), 116 - 125.
- [5] Lienhart, R., Kuranov, A., Pisarevsky, V. *Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapide Object Detection*. Microprocessor Research Lab Technical Report, May 2002.
- [6] Liu, Y., Zhang, D., Lu, G., and Ma, W.-Y. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40 (2007), 262 – 282.
- [7] Miller, G. A., Ed. WordNet: An on-line lexical database. *International Journal of Lexicography* 3, 4 (Winter 1990), 235-312.
- [8] Quack, T., Monich, U., Thiele, L. and Manjunath, B. S., Cortina: A System for Largescale, Content-based Web Image Retrieval. In *Proc. of 12th ACM Int. Conf. on Multimedia* (New York, NY, USA, 2004).
- [9] Papadopoulos, G. T., Mezaris, V., Dasiopoulou, S., Kompatsiaris, I. Semantic image analysis using a learning approach and spatial context. In *Proc. the 1st Conference on Semantic and Digital Media Technologies (SAMT 2006)* (Athens, Greece, December 6 – 9).
- [10] Popescu, A., Grefenstette, G., and Moëllic, P.-A. Using Semantic Commonsense Resources in Image Retrieval, In *Proc. of SMAP 2006* (Athens, Greece, December 4 - 5, 2006).
- [11] Smith, J. R., and S.-F. Chang, An Image and Video Search Engine for the World-Wide Web. In *Proc. of IS&T/SPIE Symposium on Electronic Imaging: Science and Technology (EI'97)* (San Jose, CA, February 1997).
- [12] Stehling, R. O., Mario. Nascimento, A., and Falcao, A.X.. A compact and efficient image retrieval approach based on border/interior pixel classification. In *Proc. of CIKM '02* (McLean, Virginia, USA, 2002).
- [13] Tiltomo system. <http://www.tiltomo.com> (consulted on January 23, 2007)
- [14] Viola, J. and Jones, M. Robust Real-time Object Detection. In *Proc. of the Second Int. Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing and Sampling*. (Vancouver, Canada, 2001).
- [15] Wang, J., Li, J., Wiedergold, G. SIMPLicity: Semantics-Sensitive Integrated Matching for Picture Libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23, 9 (September 2001), 947 – 963.
- [16] Yavlinsky, A., Heesch, D., and Rürger, S. A large scale system for searching and browsing images from the World Wide Web. In *Proc. of the Int. Conf. on Image and Video Retrieval (CIVR'06)* (Tempe, Arizona, USA, July 2006)