

# Conceptual image retrieval over the Wikipedia corpus

Adrian Popescu \*, Hervé Le Borgne, and Pierre-Alain Moëllic

CEA, LIST, Laboratoire d'ingnierie de la commnaissance multimédia et multilingue  
F-92265 Fontenay-aux-Roses, France.  
popescu\_adi@yahoo.co.uk,  
{herve.le-borgne,pierre-alain.moellic}@cea.fr,

**Abstract.** Image retrieval in large-scale databases is currently based on a textual chains matching procedure, a technique that produces good results as long as the annotations associated to pictures are accurate and detailed enough. These conditions are not met for a large majority of image corpuses, such as the Wikipedia collection, and it is interesting to explore methods that go beyond chain matching. In this paper, we present our approach to image retrieval, tested in the ImageCLEF 2008 WikipediaMM. The approach is based on a query reformulation using concepts that are semantically related to those in the initial query. For each interesting entity in the query, we used Wikipedia and WordNet to extract and list of related concepts, which were further ranked in order to propose the most salient in priority. We also made a list of visual concepts which were used in order to re-rank the answers to queries that included, implicitly or explicitly, these visual concepts. The CEA submitted two automatic runs, one based on query reformulation only and one combining query reformulation and visual concepts, which were ranked 4th and 2nd using the MAP measure..

**Key words:** We would like to encourage you to list your keywords within the abstract section

## 1 Introduction

The search for multimedia documents represents a growing trend in Web search. For this, new retrieval paradigms, going beyond text chains matching, are needed in order to better respond to user needs. Current image retrieval systems highly rely on text searching techniques and, although people search for images, the results are returned based on the associated text only. The results are obtained by simple match of the terms of the query to an index of terms associated to the images in a corpus. This technique is simple but its efficiency strongly depends on the terms associated to pictures, as well as their accuracy. For instance, in this paradigm, it would be impossible to know if a query with bridge regards the structures with this name or the cards game. If the request is relative to the

---

\* Adrian Popescu is currently...

first meaning of the term, the search engine would return only images explicitly annotated with bridge whereas it would be pertinent to propose answers for Pont Neuf or Ponte Vecchio. The exploitation of semantic structures represents a possible solution to cope with such problems if these structures are developed enough to cover the query space. In spite of flourishing research on content-based image retrieval [17], the introduction of image processing techniques in image search engines architectures is limited, for the moment, to face detection (proposed by Google, Live Search or Exalead). The adoption of more image processing techniques is conditioned by the achievement of good quality results on large amounts of data and this imperative is not met in most cases [26].

The introduction of conceptual structures and image processing techniques in image retrieval raises a number of hard questions and we try to tackle several of these questions in our work. Namely: When dealing with large conceptual domains are there enough resources available or is it necessary to enrich them? Generic semantic structures, like WordNet [7], exist and were used in image retrieval [40], [42] but they do not ensure a sufficient coverage of the query space. Wikipedia is a rich source of semi-structured and has been used to structure large quantities of knowledge [1], [25]. We exploit both WordNet and Wikipedia for extracting the information we need when processing the WikipediaMM queries. Should image processing techniques be introduced alone or should they be fused with other information? A large body of work [40], [42], [39] advocates for the use of both low level and high level image description in order to improve image search. We follow a somewhat similar approach and investigate a late fusion of textual information and low-level image description of items in the database. What terms in a query should be reformulated? When dealing with mono-conceptual queries, the answer to this query is straightforward: if knowledge about that particular concept is available, we should use it to expand the query. The problem gets complicated for more complex queries because the number of reformulations becomes rapidly unmanageable. We consider that nouns are the most important part of image queries and focus the query expansion on them. Happily, the WikipediaMM queries follow the general distribution of Web queries and contain a hefty part of mono-conceptual queries.

The remainder of this paper is structured as follows: in the next section, we describe related work; in Section 3, we present our method for automatically building conceptual structures; in Section 4, we introduce the automatic query reformulation algorithm used in the WikipediaMM task and, before concluding, we discuss the results of our approach.

## 2 Related Work

We describe related work from several relevant research areas, including: information extraction, image retrieval, query reformulation, visual concepts detection. Wikipedia is a rich resource that is used in a variety of information extraction or structuring tasks. In [35], the encyclopaedia is used to automatically construct lists of place and people names. [25] proposes a method for

cleaning the categorical tree of Wikipedia in order to obtain a sound taxonomy. The result is compared to Cyc and the precision of the results reaches 86.6%, with a recall of 89.1%. Kazama et al. [13] introduce a syntactic analysis of the first sentence in Wikipedia articles in order to extract IsA relations. These sentences are often definitional and the approach is successful in nearly 90% of the cases. [43] and [44] explore the automatic enrichment of WordNet using Wikipedia content. The authors try to extract hyponymy, hyperonymy, holonymy and meronymy relations based on lexical patterns learned from a text corpus. The overall precision of the extraction process exceeds 50% and there are a lot of incorrect relations that are extracted. DBPedia [1] is a translation of parts of Wikipedia articles to a database format, enabling structured queries over the content of the encyclopedia. The authors parse structured parts of the articles (such as info boxes, tables, or categories), which contain a fairly detailed description of the concepts described in the article and which can be later used in information retrieval tasks. The introduction of semantic structures in image retrieval architectures is a well-known practice. WordNet is exploited in a number of applications: to build a visual catalogue from the Web [42]; to propose lists of related concepts in [40]; to create multimodal similarity vectors (based on WordNet and on a visual description of the images) in [8]; to limit the conceptual neighbourhood where visually similar images are searched [26]. Wang et al. [39] reuse a taxonomy of animals created by the BBC ("BBC Science and Nature Animal Category"), which contains 620 terms. This taxonomy is enriched the description of the included concepts with visual information concerning the animal's color but also image properties (like outdoor/indoor image, photo/graph) and the result is called a multimedia ontology. They authors select 20 animal species, collect corresponding pictures from Google Image provide results for the precision at several ranks (P@20, P@40, P@60 and P@80), these values corresponding to one, two, three and four pages of results in a Web search engine. A comparison between Google Image, the use of a textual ontology and the use of a multimedia ontology and the average precision is best in the last setting, followed by the textual ontology. These results are interesting but they are limited to a specific domain, where the colour of the target concepts (animals) is stable. Image queries reformulation based on semantic resources is tested, among others, in [11] et [12]. In [11], the authors exploit ConceptNet [17] to expand image queries and report a slight improvement of results (3% for a precision around 40%) when the query expansion is used. In [12], the same group compares a WordNet based query expansion to a ConceptNet based one and conclude that the two semantic structures are complementary. The use of WordNet provides a better discrimination of the expanded queries whereas the use of ConceptNet results in more diversified queries. This finding is not surprising when considering the structure of the two resources, with ConceptNet including a larger number of inter-conceptual relations. [45] takes a different approach to query reformulation and discusses the use of query logs for expanding queries and structuring results.

Visual concepts detection has been a well studied problem in computer vision, including the work on object detection and scene recognition. It is often posed as a binary classification task consisting in deciding between the considered visual concept and other possible type of images. Hence, the general scheme of the works in this field consists of three main steps, namely region of interest detection, feature extraction and classification. Different proposals for one or several of these steps led to numerous approaches. The regions of interest of an image are the spatial localizations the feature will be extracted in the next step. The simplest approach consists of considering no region of interest and extracting the features globally, as it was the case for seminal works in image classification [9, 33]. Such a holistic representation was shown to be particularly relevant in the case of scene recognition [23, 14, 16]. Among the alternatives considered in the literature to detect ROIs, one can distinguish between the following approaches: regular [6] or random [22] sampling of patches, image segmentation [2, 34], or more often using interest point detectors [29, 19]. Once the ROI are determined, a recognition system usually extracts some visual features. Numerous approaches were proposed since the seminal work of Swain and Ballard who used global colour histograms [32], including various colour and texture descriptors [20], wavelets [30, 37] and more recently local descriptors computed around interest points. This last trend consists of computing simple features on patches around interest points [18], then aggregate them into a given number of clusters in order to define a visual vocabulary, that is former used to describe the images in term of "bag of features" [38, 10, 41, 5, 22]. An alternative to this scheme is to learn analysis filters from the learning images and define a signature from their responses [15, 16]. The last important step of visual concept detection consists of learning each concept. At this level, various classifier have been used including SVM [24, 20, 16], boosting [37], Nave Bayes [30, 5], neural networks [28], generative model [6], graphical model [21] and, regularly, K-nearest neighbours from [9] to[36].

### 3 Automatic building of conceptual structures

The acquisition of knowledge related to concepts appearing in the queries is the key element of our image retrieval method. We aim at processing diversified queries and this implies an automatic building of conceptual structures. In this section, we briefly present the employed data sources, the extraction of related concepts and the ranking of the extracted terms.

#### 3.1 Data sources

The main resource we exploited was Wikipedia. Dumps of the encyclopaedia are regularly provided for a free use . We downloaded the Mars 12, 2008 English dump, which contains over two millions articles and is provided as a single file, in XML format. Next, we split the dump into individual articles in order to process the information faster. The information in Wikipedia spans over a large number

of conceptual domains, with a high number of articles describing known people, places, entertainment, organisations animals and plants. Each article is placed in at least one category, a property that facilitates the extraction of IsA relations from Wikipedia. WordNet is a lexical database, including parts for different parts of speech, such as nouns, verbs or adjectives. As our approach focuses on nouns, we only used the nouns hierarchy which contains 117798 nominal chains, corresponding to 146312 senses. These nouns are grouped in 82115 sets of synonyms (or synsets), the sense separation being an interesting property for image retrieval tasks as it makes possible a separation of the different visual representation of the same terms. Out of the total number of synsets in WordNet, around 75% are common nouns and 25% are instances. That being, WordNet ensures an acceptable coverage for concepts but only a poor coverage of instances which have around 20000 associated synsets. For comparison, there are at least 80000 articles for person names in the English Wikipedia and around 300000 for place names .

### 3.2 Conceptual neighbourhood building

The text associated to the pictures in the WikipediaMM corpus is generally short, containing few terms. In this context, query reformulation is a way to improve recall and, if realized in a judicious way, to also improve the precision of the answers. In [27], we showed that the use of subtypes of generic concepts like dog or skyscraper is beneficial in image retrieval. This hypothesis is justified by the fact that, from a visual perspective, a concept like dog is well represented by dog races such as German shepherd, Doberman or basset. For specific concepts, having no subtypes, it is possible to build a list of synonyms in order to reformulate the queries containing them. For image queries, nouns are the most informative parts of the user's request and we focus our work on them. From the list of queries provided for WikipediaMM, we build an initial list of all nouns, which is subsequently filtered in order to eliminate visual concepts. We decided not to try to reformulate visual concepts but to employ them in the visual concepts detection framework described in [20]. The list of visual concepts contains terms like: night, day, face, portrait, graph, drawing, cartoon, photo, picture or painting. In our approach, we favour the processing of concepts in queries rather than processing words separated by blank spaces. For instance, hunting dog is regarded as a single concept and not as a composition of hunting and dog as separate terms. The same observation stands for Da Vinci paintings, three terms that form a unique concept. When searching for subtypes of hunting dog and Da Vinci paintings, the two expressions will be considered as single concepts. The first type of composed concept is a multiword and it is detected with WordNet. "Da Vinci paintings" is retained as a single concept by comparing it to the list of categories in Wikipedia. The same rule is applied for Ice hockey players or Roads in California, which are processed as single concepts. The list of concepts was first enriched using WordNet hyponyms (for those terms that, when a term existed in the hierarchy. If we get back to the example of hunting dog, the list of hyponyms contains intermediary concepts like sporting dog, hound or retriever, as well as

race names such as labrador retriever, Ibizan hound or Irish Terrier. Other terms, such as Ferrari, are not described in WordNet and they constitute an argument for using a broader resource in order to build lists of subtypes. WordNet was also used to determine the right sense for an ambiguous text queries such as the one formed of plant as text query (which is disambiguated using building, a visual concept included in the query). We were able to map this query to the second sense of plant in WordNet and extract the right set of hyponyms. A third role of WordNet was to provide a list of synonyms for terms having no subtypes in the hierarchy. For example, polar bear was enriched with ice bear, *Ursus Maritimus*, *Thalarctos Maritimus*. When using Wikipedia, we first probed each member of the list of nominal concepts (including the terms extracted from WordNet) against the categories associated to articles in the encyclopedia. Wikipedia is generally more detailed than WordNet and, when existing, the lists of subtypes extracted from WordNet were enriched with terms from the encyclopedia. For instance, Mudhol Hound or Azawakh (races of hounds) were added to the list of hyponyms of hunting dog. For categories which are not represented in WordNet, i.e. Ferrari or Da Vinci paintings, we mined Wikipedia content and extracted list of instances. Ferrari hyponyms include: Ferrari F40, Ferrari GT4 or Ferrari Testarossa. Among the extracted Da Vinci paintings, we cite: Mona Lisa, The Last Supper or The Battle of Anghiari.

### 3.3 Subtypes ranking

The lists of subtypes we constituted often contain a lot of terms (hundreds for terms like bridge or castle) and, given that they are to be used in an information retrieval application, it is necessary to rank these terms. We propose a two steps ranking process: First, the query is examined to see if it includes a useful qualifier of the concept (usually an adjective) and if so, terms associated to that qualifier are ranked first. Qualifiers appear in queries such as female players beach volleyball, red Ferrari or blue flower. These qualifiers are matched against WordNet glosses for terms in WordNet and against the categories as well as the text of the article for terms extracted from Wikipedia. Second, we use the length of the dedicated Wikipedia article to associate a pertinence value to each member of the subtypes list. The intuition behind this choice is that interesting concepts are usually described in more detail than the others. This simple way to rank entities generally gives satisfying results. For example, after ranking the list of subtypes of bridge, the first terms were: I-35w Mississippi River Bridge; San Francisco-Oakland Bay Bridge; Golden Gate Bridge; Millau Viaduct; Luding Bridge; Brooklyn Bridge and Sydney Harbour Bridge. All these terms correspond to well-known bridges. Note that, for those queries where qualifiers appear, they are priority. All ranks are normalized to values smaller than one. The term ranking will be used to order the pictures that are retrieved for a query. If a picture of the Golden Gate Bridge and one of the Luding Bridge are found for a query with bridges, they will be presented in this order on the results page.

## 4 Query reformulation and matching procedure

We performed two types of query reformulation, one implying only text and a second implying both text and visual concepts (called multimedia reformulation).

### 4.1 Textual query reformulation and matching

We preprocess the textual queries in the WikipediaMM set. First, we do not consider as concepts the stop words that appear in the queries; they are stripped off before the reformulation. For instance *bridges at night* is processed as *bridges night*. Second, nouns in the queries are stemmed and both forms of the word are searched in the text associated to the images in the dataset. Third, when possible, verbs are transformed into corresponding nouns (*dance* becomes *dancing*). Also, all capitalized letters are transformed to low-case.

The concepts are three types: Visual concepts - belonging to a pre-established closed list of terms (including *graph*, *map* or *night*). They are not subject to textual reformulation and are only used in the multimedia reformulation. Expandable concepts - terms for which we built a list of subtypes from WordNet and Wikipedia. Non-expandable concepts - terms for which there is no available reformulation. They include qualifiers (i.e. *blue*, *military*) or instances (i.e. *Golden Gate Bridge*). All concepts in a query, as well as the subtypes of expandable concepts, are tested against the text associated to images and, whenever a match is found, we increase the matching score between the query and the image. Visual concepts and expandable terms are given a higher score compared to that given to subtypes, which are, better scored than qualifiers in turn. The ranking of the image results is done by summing all the individual scores associated to query components. This ensures that the images with the most concepts in the (expanded) query appear in the descriptive text are best ranked. For instance if an image description contains both *hunting dog* and *labrador retriever*, it will be better ranked than an image described only by *hunting dog*, which, in its turn, will be better ranked than a third image described associated with *labrador retriever*. When two texts match only subtypes of the concept in the query, we use the subtypes ranking in order to differentiate between both (an image described by *labrador retriever* is ranked better than one described by *Tenterfield terrier*). In the case of the queries containing more than one concept, no image results are returned when a qualifier or a visual concept in the query matches the analyzed text.

### 4.2 Multimedia query reformulation and matching

This section describes the third layer of the system that aims at (eventually) rearranging the order of the answers returned to a query by the first two layers, depending on its content in terms of visual concepts.

We used two systems to detect visual concepts within the images. The first one is the Viola-Jones face detector that is based on the boosting of Haar wavelets

Run	MAP	P@5	P@10
ceaConTxt	0.2735	0.5467	0.4653
ceaTxt	0.2632	0.52	0.4427

**Table 1.** Main results of the CEA LIST at ImageCLEF wikipedia task

[37]. The second system (extensively described in [20]) is a set of SVM-based classifiers learnt (RBF kernel) to determine: " The type of an image: clipart, map, painting or photo. " In this last case (if the image is a photo), other sets of SVM determine whether the image is: o Indoor or outdoor o Day or night o Urban or natural scene The features extracted include texture LEP [3], colour histogram [32] and connected pixels mutual extension [31]. For each classifier, the images of the learning databases were chosen separately of the wikipedia corpus considered at imageCLEF. When more than two concepts are considered, the multi-class paradigm was solved using a one-versus-one approach. The implementation was based on LibSVM [4].

The queries were analysed to detect those which have to be filtered by one of the systems described above. Each visual concept was linked to a pre-defined list of textual concept that triggers its use. For instance, the presence of a named entity of a person (such as "Georges W Bush") will trigger the use of the face detector. The presence of the word "map" within the query will claim for the use of the "image type detector" and favour the images tagged as map while the word "cartoon" will do the same for the images classified as clipart. When a list of answers coming from the two first layers is reordered, the images detected as relevant according to the visual concept associated to the query are put at the head of the list without changing their relative order.

## 5 Results and discussion

Table 1 gives the main results of the two runs we submitted. The run ceaTxt is the output of the textual query reformulation and matching only, while the run ceaConTxt is the output of the full system including the multimedia query reformulation and matching. The results are given in terms of Mean Average Precision as well as the precision at ranks five and ten.

Our system returns good results, which were ranked at the fourth and second place of the WikipediaMM task at imageCLEF 2008. The difference between our two runs shows the slight interest of the multimedia reformulation and re-arrangement that led to an improvement of one point in terms of MAP (from 0.263 to 0.273). It is worth noting that about half of the images were judged as relevant among the ten first answers returned by our system, demonstrating a practical interest for a real user.

## 6 Conclusions and perspectives

We proposed a new scheme to exploit both textual and visual information in the context of image retrieval. The approach is based on a query reformulation using concepts that are semantically related to those in the initial query. For each interesting entity in the query, we used Wikipedia and WordNet to extract and list of related concepts, which were further ranked in order to propose the most salient in priority. These answers were ultimately rearranged in function of the query reformulation in terms of visual concepts. The results submitted at ImageCLEF 2008 were ranked 4th and 2nd with a mean average precision of 0.2632 and 0.2735. The small difference between the two submitted runs shows that the greater contribution to the final results was probably due to the use of conceptual structures, although a rigorous comparison would have required submitting a run with the third layer (visual concept detection) only. Nevertheless, the improvement of the results' precision accounts for the interest of introducing visual concept detection in the retrieval schema. We described an ongoing work and a number of features of our system are currently under investigation. The detection of associated concepts is currently limited to the use of Wikipedia and WordNet. We plan on extending our approach so as to exploit search engine snippets, in order to improve the coverage of the resources. Also, there exist domain-related semantic structures, like Geonames for geography, which could be used to improve the coverage. A second line of work concerns the concept ranking process described in this paper. While simple and generally effective, the current approach can certainly be ameliorated if, for instance, we favour unambiguous hyponyms over ambiguous ones. Concerning the third layer, we are currently exploring a finer grained filtering of visual concepts. Indeed, the current implementation uses a classifier that does not require any setting put aside the choice of the kernel. Hence, stepping up our system (in terms of number of visual concepts) is a problematic subject on its own that consists of being able to build the learning databases (images and triggering words) automatically for a large number of concepts.

## 7 BibTeX Entries

The correct BibTeX entries for the Lecture Notes in Computer Science volumes can be found at the following Website shortly after the publication of the book: <http://www.informatik.uni-trier.de/~ley/db/journals/lncs.html>

**Acknowledgments.** The heading should be treated as a subsection heading and should not be assigned a number.

## 8 The References Section

In order to permit cross referencing within LNCS-Online, and eventually between different publishers and their online databases, LNCS will, from now on, be

standardizing the format of the references. This new feature will increase the visibility of publications and facilitate academic research considerably. Please base your references on the examples below. References that don't adhere to this style will be reformatted by Springer. You should therefore check your references thoroughly when you receive the final pdf of your paper. The reference section must be complete. You may not omit references. Instructions as to where to find a fuller version of the references are not permissible.

We only accept references written using the latin alphabet. If the title of the book you are referring to is in Russian or Chinese, then please write (in Russian) or (in Chinese) at the end of the transcript or translation of the title.

The following section shows a sample reference list with entries for journal articles [44], an LNCS chapter [?], a book [?], proceedings without editors [?] and [?], as well as a URL [?]. Please note that proceedings published in LNCS are not cited with their full titles, but with their acronyms!

## References

1. S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives. "Dbpedia : A nucleus for a web of open data". In Proceedings of the 6th International Semantic Web Conference (ISWC), Volume 4825 of Lecture Notes in Computer Science, pages 722-735. Springer, (2008).
2. K. Barnard, P. Duygulu, D. A. Forsyth, Nando de Freitas, D.M. Blei, M.I. Jordan: Matching Words and Pictures. *Journal of Machine Learning Research* 3: 1107-1135 (2003)
3. Y.-C. Cheng and S.-Y. Chen. Image classification using color, texture and regions. *image and Vision Computing*, 21 :759-776, 2003.
4. Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
5. Deselaers T., Keysers D., Ney H., "Discriminative Training for Object Recognition using Image Patches", *CVPR*, vol. 2, San Diego, CA, USA, IEEE, pp. 157-162, 2005
6. L. Fei-Fei and P. Perona. A Bayesian Hierarchical Model for Learning Natural Scene Categories. *IEEE Comp. Vis. Patt. Recog.* 2005
7. C. Fellbaum, editor. *WordNet : an electronic lexical database*. MIT Press, (1998).
8. M. Ferecatu, N. Boujemaa, M. Crucianu. "Semantic interactive image retrieval combining visual and conceptual content description", *Multimedia Syst.*, 13(5-6), pp. 309-322, (2008).
9. M. M. Gorkani and R. W. Picard, Texture Orientation for Sorting Photos at a Glance, *Proc. ICPR*, Oct 1994, TR #292.
10. K. Grauman and T. Darrell. Pyramid match kernels: Discriminative classification with sets of image features. In *Proc. ICCV*, 2005
11. M.-H. Hsu, H.-H. Chen. "Information retrieval with commonsense knowledge". In *SIGIR '06 : Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 651-652, New York, NY, USA, (2006). ACM.
12. M.-H. Hsu, M.-F. Tsai, H.-H. Chen. "Query expansion with ConceptNet and Wordnet : An intrinsic comparison". In *Proceedings of the Third Asia Information Retrieval Symposium Information Retrieval Technology*, pages 1-13, (2006).

13. J. Kazama, K. Torisawa. "Exploiting wikipedia as external knowledge for named entity recognition". In Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pages 698-707, (2007). [14] H. Le Borgne and A. Gurin-Dugu. Sparse-Dispersed Coding and Images Discrimination with Independent Component Analysis, Third International Conference on Independent Component Analysis and Signal Separation, San Diego, California, 9-13, 2001
14. H. Le Borgne, A. Gurin-Dugu, A. Antoniadis. Representation of images for classification with independent features. *Pattern Recognition Letters*, 25(2):141-154, 2004
15. H. Le Borgne, A. Gurin-Dugu, N.E. O'Connor, Learning Mid-level Image Features for Natural Scene and Texture Classification, *IEEE transaction on Circuits and Systems for Video Technology*, 17(3):286-297, march 2007.
16. Y. Liu, D. Zhang, G. Lu, W.-Y. Ma. "A survey of content-based image retrieval with high-level semantics", *Pattern Recogn.*, 40(1), pp. 262-282, (2007).
17. D. Lowe. Object Recognition from Local Scale-Invariant Features. In Proceedings of the International Conference on Computer Vision, pages 1150-1157, Corfu, Greece, September 1999.
18. K. Mikolajczyk and C. Schmid. Indexing Based on Scale Invariant Interest Points. In Proceedings of the International Conference on Computer Vision, pages 525-531, 2001.
19. C. Millet, Automatic image annotation: consistent annotation, and creating automatically a learning database. PhD thesis, 2007.
20. P. Murphy, A. Torralba and W. T. Freeman. Using the forest to see the trees: a graphical model relating features, objects and scenes. *Adv. in Neural Information Processing Systems 16 (NIPS)*, Vancouver, BC, MIT Press, 2003.
21. E. Novak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. In *IEEE European Conf. on Computer Vision*, 2006
22. A. Oliva and A. Torralba. Modelling the shape of a scene: a holistic representation of the spatial envelope, *International Journal of Computer Vision*, 42(3):145-175, 2001.
23. C. Papageorgiou and T. Poggio. A trainable system for object detection. *Intl. J. Computer Vision*, 38(1):15-33, 2000.
24. S. P. Ponzetto, M. Strube. "Deriving a large scale taxonomy from wikipedia". In Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, (2007).
25. A. Popescu, C. Millet, P-A. Mollic Ontology Driven Content Based Image Retrieval, *CIVR 2007 - posters session*, July 9 - 11, 2007, Amsterdam, The Netherlands.
26. A. Popescu, P-A. Mollic, I. Kanellos A Conceptual Approach to Web Image Retrieval , *LREC 2008*, May 28 - 30, 2008, Marrakech, Morocco.
27. H.A. Rowley, S. Baluja, T. Kanade, "Rotation Invariant Neural Network-Based Face Detection," *cvpr*, p. 963, 1998 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'98)*, 1998
28. C. Schmid and R. Mohr. Local Grayvalue Invariants for Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530-535, May 1997.
29. H. Schneiderman and T. Kanade. A statistical model for 3D object detection applied to faces and cars. *Conference on Computer Vision and Pattern Recognition*, 2000.

30. R. O. Stehling , M. A. Nascimento , A. X. Falco. A compact and efficient image retrieval approach based on border/interior pixel classification. Proceedings of the eleventh international conference on Information and knowledge management, pp 102-109, McLean, Virginia, USA 2002.
31. M. Swain and D. Ballard, Color Indexing, International Journal of Computer Vision, 7(1):11-32, 1991.
32. M. Szummer and R. W. Picard, "Indoor-outdoor image classification," Int. Workshop on Content-based Access of Image and Video Databases, pp. 42-51, Jan. 1998.
33. S. Tollari, H. Glotin, J. Le Maitre: Enhancement of Textual Images Classification Using Segmented Visual Contents for Image Search Engine. Multimedia Tools and Applications 25(3): 405-417 (2005)
34. A. Toral, R. Muoz. "A proposal to automatically build and maintain gazetteers for named entity recognition by using wikipedia". In NEW TEXT - Wikis and blogs and other dynamic text sources, Trento, (2006).
35. A. Torralba, R. Fergus, W. Freeman, "80 million tiny images: a large dataset for non-parametric object and scene recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, 20 May 2008
36. P. Viola and M.J. Jones. Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001.
37. C. Wallraven, B. Caputo, and A. Graf. Recognition with local features: the kernel recipe. In Proc. ICCV, volume 1, pages 257-264, 2003.
38. H. Wang, S. Liu, L.-T. Chia. "Does ontology help in image retrieval ? - a comparison between keyword, text ontology and multi-modality ontology approaches". In MULTIMEDIA '06 : Proceedings of the 14th annual ACM international conference on Multimedia, pages 109-112, New York, NY, USA, (2006). ACM.
39. J. Yang, L. Wenyin, H. Zhang, Y. Zhuang. "Thesaurus-aided approach for image browsing and retrieval", Proceedings of ICME 2001, (2001).
40. J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: An in-depth study. Technical Report RR-5737, INRIA Rhne-Alpes, 2005.
41. X. J.Wang, W. Y. Ma, X. Li. "Data-driven approach for bridging the cognitive gap in image retrieval". In Proceedings of the 2004 IEEE International Conference on Multimedia and Expo, Volume 3, pages 2231-2234, Taipei, Taiwan, (June 2004).IEEE.
42. M. Ruiz-Casado, E. Alfonseca, P. Castells. "Automatic assignment of wikipedia encyclopedic entries to wordnet synsets", Advances in Web Intelligence, pages 380-386, (2005).
43. M. Ruiz-Casado, E. Alfonseca, P. Castells. Automatising the learning of lexical patterns : An application to the enrichment of wordnet by extracting semantic relationships from wikipedia", Data Knowl. Eng., 61(3), pp. 484-499, (2007).
44. S. P. Liao, P. J. Cheng, R. C. Chen, L. F. Chien. Liveimage : Organizing web images by relevant concepts". In Proc. of the Workshop on the Science of the Artificial 2004, pages 210:220, (2005)