# Knowledge-based Expansion for Image Indexing

R.C.F. Wong and C.H.C. Leung
Dept. of Computer Science, Hong Kong Baptist University,
{cfwong,clement}@comp.hkbu.edu.hk

***Abstract-*** *The number of images available in online repositories is growing dramatically and image retrieval has been become one of the most popular activities on Internet. The effectiveness of image retrieval depends on meaningful indexing. In this paper, we propose an extension of image annotation models which are using knowledge-based expansion and contextual feature-based index expansion. Our system is evaluated quantitatively using more than 100,000 web images and around 1,000,000 tags. Experimental results indicate that this approach is able to deliver highly competent performance.*

## 1 Introduction

Since the past decade, image retrieval has been become one of the most popular activities on Internet. As the number of images available in online repositories is growing dramatically, exploring the frontier between image and language is an interesting and challenging task. Research in image retrieval has reflected the dichotomy inherent in the semantic gap, and is divided between two main categories: concept-based image retrieval and content-based image retrieval. The former focuses on retrieval by objects and high-level concepts, while the latter focuses on the low-level visual features of the image.

Low-level visual features are indicated by visual content descriptors in order to support users in accessing the knowledge embedded in images. These methods aim at capturing image similarity by relying on some specific characteristic of images; typically, these models are based on color, texture and shape [4, 6, 9, 11, 14, 23, 30, 32]. As discussed in [31], in order to compute these descriptors, the image often has to be segmented into parts, which aims to determine image objects. Current methods of image segmentation include [8, 12, 16, 17, 24, 29]: partitions, sign detection, region segmentation. They compute general similarity between images based on statistical image properties [1–3, 18, 22, 26, 27]. Some studies [12, 20] include users in a search loop with a relevance feedback mechanism to adapt the search parameters based on user feedback. Semantic annotation of the image database combined with a region based image decomposition is used, which aims to extract semantic properties of images based on spatial distribution of color and texture properties [9,11,14,15,23,30,32]. However, an advantage of using low-level features is that, unlike high-level concepts, they do not incur any indexing cost as they can be extracted by automatic algorithms. In contrast, direct extraction of high-level semantic content automatically is beyond the capability of current technology. Some research [5] focuses on implicit image annotation which involve an implicit and, in consequence, augments the original indexes with additional concepts that are related to the query.

With the advent of Semantic Web technology, knowledge is playing a key role as the core element of knowledge representation architecture in Semantic Web. Some effort [7, 13, 19, 25, 28] has been made for image retrieval using Semantic Web techniques.

Our work is related to generative modelling approaches. In [31], a semantic annotation technique named Automatic Semantic Annotation (ASA) approach is developed which is based on the use of image parametric dimensions and metadata . Using decision trees and rule induction, a rule-based approach to formulate explicit annotations for images fully automatically is developed, so that, semantic query such as ”sunset by the sea in autumn in New York” can be answered and indexed purely by machine. In this paper, we propose an extension of such image annotation models by using knowledge-based expansion and contextual feature-based index expansion. Our system is evaluated quantitatively using more than 100,000 web images and over 990,000 tags. Experimental results indicate that this approach is able to deliver highly competent performance.

## 2 Expansion algorithms

### 2.1 Knowledge-based expansion

In certain applications, the presence of particular objects in an image often implies the occurrence of other objects. The application of such inferences will allow the index of an image to be automatically expanded.

Aggregation hierarchical expansion is a particularly useful technique, which relates to the aggregation hierarchy of sub-objects that constitute an object. The objects can be classified as:

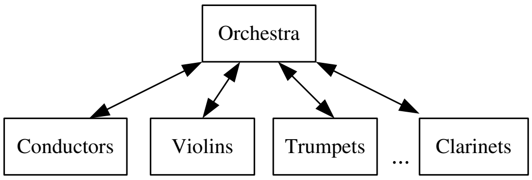


**Figure 1. Hierarchical expansion**

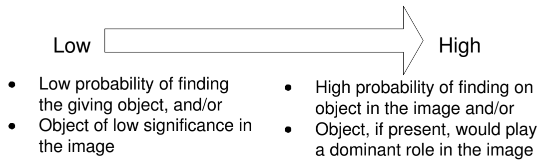


**Figure 2. Annotation measure**

(a) concrete, where the relevant objects are well-defined (e.g. an ”orchestra” expanded to violins, trumpets, clarinets etc, see Fig. 1)

(b) abstract, where the objects are not concretely defined (e.g. although ”conflict” is not a definite visual object, it contains certain common characteristics).

### 2.2 Contextual feature-based index expansion

Here, we shall establish associations between low-level features with high-level concepts, and such associations will take the following forms.

The presence of certain low-level features $F$ may suggest a finite number of $m$ object possibilities. Sometimes, a combination of basic features may be used to infer the presence of high-level concepts for inclusion in the semantic index.

Basic features alone may not be sufficient to infer the presence of specific objects, but such features if augmented by additional information may lead to meaningful inferences. When a particularly context is known, a concept may be indexed more precisely. Such contextual information will typically be provided through knowledge-based expansion, which may lead to the creation of a new index term, or a revision of the score of an existing index term. This will give rise to an iterative feedback loop where the determination of new objects will lead to new meaningful feature-object combinations, where further objects may be determined.

### 2.3 Annotation measure

The reliability of a given annotation will given rise to a numerical measure, which signifies how good the annotation is. For annotations with a low measure, this would mean that the annotation is not very occurrence, or in extreme cases, what is being annotated is absent from an image. A high annotation measure indicates that the chance of finding the corresponding object or content in the given image is high. In addition, apart from measuring the likelihood of of whether something is present or not, it can be used to indicate the importance of an object in the image. For example, an object which is very prominently present in the foreground of an image would have a much large value than an object of small size in the remote background. Hence the annotation measure is used to signify two aspects.

(a) the likelihood of finding the object in the given image

(b) the prominence of the object in the given image (Fig. 2).

## 3 Experimental evaluation

Our main purpose in introducing knowledge-based expansion into the image retrieval problem and using the subobjects as surrogate terms for general queries is to improve the precision in the image sets. In this evaluation, we mainly focus on the knowledge-based expansion and contextual feature-based index expansion.

The index elements are organized and used to build the basic content index within a relational database. The relational database is designed for maximum query effectiveness by distributing the semantic elements across different relations. A further index is built on top of these relations to support rapid discovery.

The effectiveness of our approach is evaluated experimentally. A set of standard evaluation queries are used for experimentation. Comparison is made between base-level indexing and the expanded level indexing, and the widely accepted measures of retrieval performance of precision and recall are used to assess system performance.

To numerically assess the accuracy and effectiveness of our annotation approach, we have retrieved 103,521 sets of images with 991,074 associated tags from flickr.com which are a popular photo sharing web site and online community platform offering a fairly comprehensive web-service API that allows developers to create applications that can perform almost any function on images.

The quality assessment of the machine-inferred boundaries between parts of the depicted scenes is based on the precision. In our evaluation, we decide that a relevant image must include a representation of the category in such a manner that a human should be able to immediately associate it with the assessed concept.

### 3.1 Results

In relation to image acquisition, many images may be broken down to few basic scenes, such as nature and wildlife, portrait, landscape and sports. In the case of aggregation hierarchical expansion, we decided to test our system using the aggregation hierarchy of basic categories ”night scenes” and extend the image hierarchy to find a sub-scene

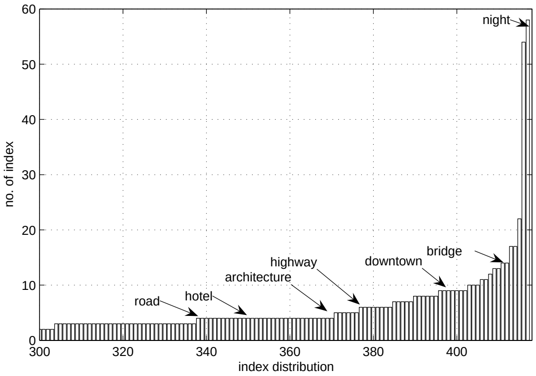


**Figure 3. Index distribution associated with night scene images**

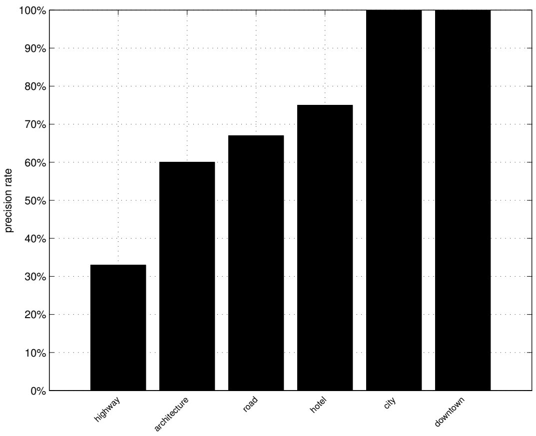


**Figure 4. Experimental results on aggregation hierarchical expansion**

"night scene of downtown", "downtown" can be expanded to "business district", "commercial district", "city center" and "city district", while "city district" can be expanded to "road", "building", "architecture", "highway" and "hotel".

In [31] by using decision trees and rule induction, a rule-based approach to formulate explicit annotations for images fully automatically has been developed. To extend the approach, firstly, we annotate night scenes based on the prior rule-based approach to extract 422 out of 103,527 images. We also gather 1108 tags associated with those images and totally 417 unique terms are formed. We list the top 117 out of 417 unique terms list in Fig. 3. We present the results of the evaluation in Fig. 4.

To establish associations between low-level features with high-level concepts, associating basic features with semantic concepts may be applied to arbitrary images for inclusion

**Figure 5. Contextual feature-based index expansion with edge detection algorithms**

in the semantic index. Edge detection is a methodology in image processing and computer vision, particularly in the areas of feature detection and feature extraction, to refer to algorithms which aim at identifying points in a digital image at which the image brightness changes sharply or more formally has discontinuities. Here, we adapt edge detection algorithms [10, 21] to extract high-level concepts from low-level features.

From [10], the framework near-circular Gaussian-based image derivative operators have been developed via the use of a virtual mesh and are proven to reduce angular error when detecting edges over a range of orientations. The edge detection operators are based on first and second derivative approximations, corresponding to a first directional derivative $\partial u/\partial b \equiv \underline{b} \cdot \underline{\nabla u}$ and a second directional derivative $-\underline{\nabla} \cdot (B\underline{\nabla u})$ , and are defined by the functionals [10]

$$E_i^\delta(U) = \int_\Omega \underline{b_i} \cdot \underline{\nabla} U \zeta_i^\delta d\Omega \tag{1}$$

and

$$Z_i^\delta(U) = \int_\Omega \underline{\nabla} U \cdot (B_i \underline{\nabla} U \zeta_i^\delta) d\Omega \tag{2}$$

Here $B = \underline{b}\, \underline{b}^T$ and $b = (cos\theta, sin\theta)$ is the unit direc-

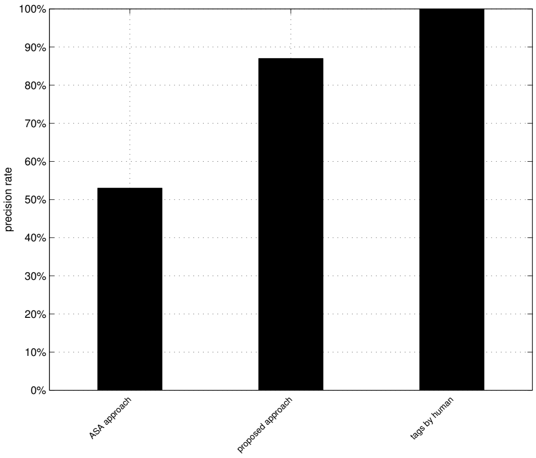


**Figure 6. Experimental results on contextual feature-based index expansion**

tion. The special case of the Laplacian operator is represented by $Z_i^\delta$ with $B$ taken to be the identity matrix [10].

Here, we select one "downtown" image manually from the image set. We have carried out evaluation (shown in Fig. 6 by comparing the original Automatic Semantic Annotation (ASA) approach) with our approach which combines the original ASA approach with vertical edge detection algorithms (see Fig. 5) and the use of human tags. Our experiment indicates that tags by human deliver excellent precision rate with 100% precision but this tagging approach relies heavily on human involvement. For the ASA Approach combining with edge detection algorithms, the precision rate grows to 87.1%. Clearly, compared to annotation without the contextual feature-based index expansion enabled, the performance is around 52.8%. From the joint application of these, we can formulate semantic annotations for specific image fully automatically and index images purely by machine without any human involvement.

## 4 Conclusion and future works

In this paper, we propose an extension of image annotation models which uses knowledge-based expansion and contextual feature-based index expansion. Our system is evaluated quantitatively, and experimental results indicate that this approach is able to deliver highly competent performance. Our approach, not only demonstrates the applicability of knowledge-based expansion to the image annotation problem, but also using the sub-objects as surrogate terms for general queries is to improve the precision in the image sets.


[1] G. Amato and C. Meghini. Combining features for image retrieval by concept lattice querying and navigation. *International Conference on Image Analysis and Processing Workshops*, 0:107–112, 2007.

[2] J. Amores, N. Sebe, and P. Radeva. Context-based object-class recognition and retrieval by generalised correlograms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10):1818–1833, October 2007.

[3] V. Athitsos, J. Alon, S. Sclaroff, and G. Kollios. Boostmap: A method for efficient approximate similarity rankings. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 02:268–275, 2004.

[4] I. Azzam, A. G. Charlapally, C. H. C. Leung, and J. F. Horwood. Content-based image indexing and retrieval with xml representations. *Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong*, pages 181–185, 2004.

[5] I. A. Azzam, C. H. C. Leung, and J. F. Horwood. Implicit concept-based image indexing and retrieval. In *Proceedings of the IEEE International Conference on Multi-media Modeling*, pages 354–359, Brisbane, Australia, January 2004.

[6] I. A. Azzam, C. H. C. Leung, and J. F. Horwood. A fuzzy expert system for concept-based image indexing and retrieval. *International Conference on Multimedia Modeling Conference*, 0:452–457, 2005.

[7] K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135, 2003.

[8] Y. Chen, J. Z. Wang, and R. Krovetz. Content-based image retrieval by clustering. In *MIR'03: Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, pages 193–200, New York, NY, USA, 2003. ACM.

[9] J. S. Cho and J. Choi. Contour-based partial object recognition using symmetry in image databases. In *SAC'05: Proceedings of the 2005 ACM symposium on Applied computing*, pages 1190–1194, New York, NY, USA, 2005. ACM Press.

[10] S. Coleman, B. Scotney, and D. Kerr. Integrated edge and corner detection. *International Conference on Image Analysis and Processing*, 0:653–658, 2007.

[11] D. Cremers, M. Rousson, and R. Deriche. A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *International Journal of Computer Vision*, 72(2):195–215, 2007.

[12] R. Datta, J. Li, and J. Z. Wang. Content-based image retrieval: approaches and trends of the new age. In *MIR'05: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*, pages 253–262, New York, NY, USA, 2005. ACM.

[13] L. Fan and B. Li. A hybrid model of image retrieval based on ontology technology and probabilistic ranking. *Web Intelligence*, 0:477–480, 2006.

[14] J. Gausemeier, J. Fruend, C. Matysczok, B. Bruederlin, and D. Beier. Development of a real time image based object recognition method for mobile ar-devices. In *International conference on Computer graphics, virtual Reality, visualisation and interaction in Africa*, pages 133–139, New York, NY, USA, 2003. ACM Press.

[15] K. Hornsby. Retrieving event-based semantics from images. *International Symposium on Multimedia Software Engineering*, 00:529–536, 2004.

[16] M. Jian, J. Dong, and R. Tang. Combining color, texture and region with objects of user's interest for content-based image retrieval. *International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, 01:764–769, 2007.

[17] R. Krishnapuram, S. Medasani, S. H. Jung, Y. S. Choi, and R. Balasubramaniam. Content-based image retrieval based on a fuzzy approach. *IEEE Transactions on Knowledge and Data Engineering*, 16(10):1185–1199, 2004.

[18] J. Li and J. Z. Wang. Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1075–1088, 2003.

[19] H. Lieberman and H. Liu. Adaptive linking between text and photos using common sense reasoning. *Adaptive Hypermedia and Adaptive Web-Based Systems, Second International Conference, AH 2002, Malaga, Spain*, pages 2–11, May 2002.

[20] D. Liu and T. Chen. Content-free image retrieval using bayesian product rule. *IEEE International Conference on Multimedia and Expo*, 0:89–92, 2006.

[21] J. Manikandan, B. Venkataramani, and M. Jayachandran. Evaluation of edge detection techniques towards implementation of automatic target recognition. *International Conference on Computational Intelligence and Multimedia Applications*, 2:441–445, 2007.

[22] A. P. Natsev, A. Haubold, J. Tešić, L. Xie, and R. Yan. Semantic concept-based query expansion and re-ranking for multimedia retrieval. In *International conference on Multimedia*, pages 991–1000, New York, NY, USA, 2007. ACM.

[23] R. Pawlicki, I. Kókai, J. Finger, R. Smith, and T. Vetter. Navigating in a shape space of registered models. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1552–1559, 2007.

[24] A. Perina, M. Cristani, and V. Murino. Natural scenes categorization by hierarchical extraction of typicality patterns. *International Conference on Image Analysis and Processing*, pages 801–806, 2007.

[25] A. Popescu, G. Grefenstette, and P. A. Moellic. Using semantic commonsense resources in image retrieval. In *International Workshop on Semantic Media Adaptation and Personalization*, pages 31–36, Washington, DC, USA, 2006. IEEE Computer Society.

[26] K. Stevenson and C. H. C. Leung. Comparative evaluation of web image search engines for multimedia applications. *IEEE International Conference on Multimedia and Expo*, 0:4 pp., 2005.

[27] Y. Sun, S. Shimada, and M. Morimoto. Visual pattern discovery using web images. *MIR'06: Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, pages 127–136, 2006.

[28] A. M. Tam and C. H. C. Leung. Semantic content retrieval and structured annotation: Beyond keywords. *ISO/IEC JTC1/SC29/WG11 MPEG00/M5738, Noordwijkerhout, Netherlands*, March 2000.

[29] N. Vasconcelos. From pixels to semantic spaces: Advances in content-based image retrieval. *IEEE Computer*, 40(7):20–26, 2007.

[30] J. Vogel, A. Schwaninger, C. Wallraven, and H. H. Bülthoff. Categorization of natural scenes: Local versus global information and the role of color. *ACM Transactions on Applied Perception*, 4(3):19, 2007.

[31] R. C. F. Wong and C. H. C. Leung. Automatic semantic annotation of real-world web images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):1933–1944, November 2008.

[32] T. Zöller and J. M. Buhmann. Robust image segmentation using resampling and shape constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7):1147–1164, July 2007.