A Semantic Vector Space for Query by Image Example

João Magalhães[1], Simon Overell[1], Stefan Rüger[1,2]

[1]Department of Computing
Imperial College London
South Kensington Campus
London SW7 2AZ, UK

[2]Knowledge Media Institute
The Open University
Walton Hall
Milton Keynes MK7 6AA, UK

(j.magalhaes@imperial.ac.uk, simon.overell01@imperial.ac.uk, s.rueger@open.ac.uk)

ABSTRACT

Content-based image retrieval enables the user to search a database for visually similar images. In these scenarios, the user submits an example that is compared to the images in the database by their low-level characteristics such as colour, texture and shape. While visual similarity is essential for a vast number of applications, there are cases where a user needs to search for semantically similar images. For example, the user might want to find all images depicting bears on a river. This might be quite difficult using only low-level features, but using concept detectors for "bear" and "river" will produce results that are *semantically* closer to what the user requested. Following this idea, this paper studies a novel paradigm: query by semantic multimedia example. In this setting the user's query is processed at a semantic level: a vector of concept probabilities is inferred for each image and a similarity metric computes the distance between the concept vector of the query and of the concept vectors of the images in database. The system is evaluated with a COREL Stock Photo collection.

Categories and Subject Descriptors

H.3.1 [**Content Analysis and Indexing**]: Abstracting methods.

General Terms

Algorithms, Measurement, Experimentation.

Keywords

Semantic multimedia retrieval, query by semantic example.

1. INTRODUCTION

Information retrieval systems have always forced humans to describe their query in terms of a written language. While in text retrieval we express our query in the same format as the document (text), in multimedia retrieval system this is more difficult due to semantic ambiguities. The user is not aware of the low-level representation of documents, e.g. colour, texture, or shape features. Thus a multimedia information-retrieval system relies on algorithms that model semantic concepts in terms of low-level features. However, by forcing a user's idea to go trough this abstraction process (translating an idea into words) part of the original idea may be filtered leading to a less expressive query. This issue is even more visible in multimedia retrieval systems since multimedia information is very rich in terms of information content and expressiveness.

SIGIR'07, July 23-27, 2007, Amsterdam, The Netherlands.

Early research in this area produced systems where the user would draw a sketch of what he wanted to search for. QBIC [6] is a well known systems of this type. Such systems work well when one wants to search for images that are visually *very* similar to the sketch image. Taking one step further relevance feedback systems use the user input to compose a set of visually positive and negative examples that are different representations of the same semantic request. The system is still not aware of any semantics as it represents images by their low-level features.

Systems that are aware of multimedia semantics have already flourished in the multimedia information-retrieval community. These systems allow the user to specify a set of keywords or concepts, which are thus used to search for multimedia content containing those concepts. This is already a big step from previous approaches towards more semantic search engines but in some cases (if not most cases) it still may be too limiting: semantic multimedia content captures knowledge that goes beyond a limited set of concepts.

These types of approaches can produce good results but it puts an extra burden on the user that now has to formalize its "creative idea". So far, the user had three ways to express his idea: by depicting the idea visually (visual sketch), by providing several positive and negative examples (relevance feedback), or by expressing the idea textually (image annotation). In all cases it might be restraining the user in terms of creativity or expressiveness. Thus, the user should be able to formulate a query with a "semantic example" of what he wants to obtain. Of course the example is not semantic per se but the system will look at its semantic content and not at its low-level characteristics (e.g. colour or texture). This means that the system will infer the semantics of an image and use those semantics to search an image database that has been indexed with the same semantic analysis algorithm.

Note the clear distinction between the paradigm followed on this paper and the previous paradigms: user queries are not processed at the visual features level but at the semantic level. Thus, instead of matching images using their visual feature vectors we match images with their concept feature vector. This paper contributes to the study and evolution of this new paradigm.

The paper is organised as follows: Section 2 presents related work; Section 3 presents the query-by-semantic-example system; Section 4 describes the algorithm that infers the semantics of multimedia (both for query and database); Section 5 presents the semantic similarity metric; the experiments and the corresponding discussion are presented in Section 6 and Section 7, respectively. Finally, conclusions are presented in Section 8.

2. RELATED WORK

The problem of query by semantic multimedia example can be divided into two parts: (1) the multimedia analysis part that links keywords (or concepts) to multimedia, and (2) the semantic distance between the user examples and the existing elements of the multimedia database.

The initial annotation of multimedia content with concepts can be done manually, with some automatic algorithm, or with some set of heuristics. Only in very limited circumstances can manual methods provide a real solution. Automatic algorithms are by far the most attractive ones involving a low analysis cost when compared to the manual method. Automatic algorithms are all based on some statistical modelling technique of low-level features. Several techniques to model a concept with different types of probability density distributions have been proposed: Feng and Manmatha [5] proposed a Bernoulli model with a vocabulary of visual terms for each keyword, Yavlinsky et al. [22] deployed a nonparametric distribution, Carneiro and Vasconcelos [1] a semi-parametric density estimation, while Magalhães and Rüger [15] engaged a maximum entropy framework. The above methods use features extracted from the multimedia itself, but heuristic techniques rely on metadata attached to the multimedia: for example, Lu et al [13] analyse HTML text surrounding an image and assign the most relevant keywords to an image. We follow the maximum entropy approach by Magalhães and Rüger [15].

The same semantic analysis is applied to both the multimedia database and to the example. The second step in the problem is to explore the semantic similarity between the user's examples and the multimedia documents. These links can be computed automatically (pure query by semantic example) or semi-automatically (relevance feedback).

In most relevance feedback literature these links are initialised with some predefined set of weights and an iterative algorithm updates the weights of these relations based on the feedback from the different users. Relevance feedback iteratively re-ranks results according to the positive and negative semantic examples successively specified by the user. Yang et al. [21] implemented a relevance feedback algorithm that works on a semantic space created from image clusters that are labelled with the most frequent concept on that cluster. Semantic similarity is then computed between the examples and the image clusters. Lu et al [13] proposed a relevance feedback system that labels images with the previously described heuristic and updates these semantic relations according to the user feedback. The semantic links between the examples and the keywords are heuristically updated/removed. Zhang and Chen [23] followed an active learning approach, and He et al [7] applied spectral methods to learn the semantic space from the user's feedback. Other relevance feedback approaches have been proposed by Zhou and Huang [25], Chang et al. [2], and Wang and Li [20], and a good overview is Heesh and Rüger [8].

Moving away from semi-automatic retrieval, Rasiwasia [18] proposed an automated retrieval framework that computes the semantic similarity with a distance metric to rank images according to the semantics of the given query. They start by extracting semantics with an algorithm based on a hierarchy of mixtures [1]. Next, they compute the semantic similarity as the Kullback-Leibler divergence, and evaluate the system with the traditional precision measure by considering an image relevant if it shares one or more concepts with the query. This evaluation methodology does not account for the fact that one image sharing two concepts with the query is probably more relevant than an image sharing just one concept. We will discuss this fact and propose the use of rank correlation to evaluate query-by-semantic-example systems. Our system is conceptually similar to the one proposed by Rasiwasia [18] but with a different semantic multimedia analyser and semantic multimedia similarity.

3. QUERY BY SEMANTIC EXAMPLE

The implemented query by semantic example system is divided into three parts: the semantic multimedia analyser, the indexer, and the semantic multimedia retrieval. Figure 1 presents the architecture of the system.

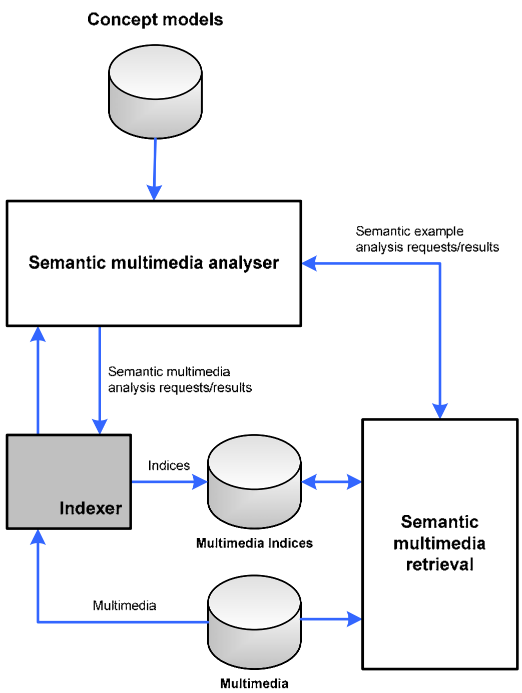

Figure 1 – Query by semantic example system.

In automatic retrieval systems, processing time is a pressing feature that directly impacts the usability of the system. We envisage a responsive system that processes a query and retrieves results within 1 second per user, meaning that to support multiple users it must be much less than 1 second.

Semantic Multimedia Analyser

The semantic multimedia analyser infers the concepts probabilities and is designed to work in less than 100 ms. Another important issue is that it should also support a large number of keywords so that the semantic space can accommodate the semantic understanding that the user gives to the query. Section 4 presents the semantic multimedia analyser used in this paper, see [15] for details.

Indexer

Indexer uses a simple storage mechanism capable of storing and providing easy access to each concept of a given multimedia document. It is not optimised for time complexity. The same indexing mechanisms used for content based image retrieval can be used to index content by semantics. This topic is outside the scope of this paper.

Semantic Multimedia Retrieval

The final part of the system is the semantic multimedia retrieval in charge of retrieving the documents that are semantically close to the given query. First it must run the semantic multimedia analyser on the example to obtain the concept vector of the query. Then it searches the database for the relevant documents according to a semantic similarity metric on the semantic space of concepts. In this part of the system we are only concerned with studying functions that mirror human understanding of semantic similarity. Section 5 will detail the implemented semantic similarity metric.

4. SEMANTIC MULTIMEDIA ANALYSER

In this section we will see how to model concepts in terms of feature data of images. Following the approach proposed by Magalhães and Rüger [15], each concept w_i is represented by a maximum entropy model,

$$p(w_i \mid V) = \mathrm{logit}(\beta^{w_i}\,\mathrm{F}(V)),$$

where $\mathrm{F}(V)$ is a transformation of visual feature vector V, and β^{w_i} is the vector of the regression coefficients for concept w_i. Next we will present visual feature data transformation $\mathrm{F}(V)$ and the implemented maximum entropy model.

4.1 Feature Data Representation

We create a visual vocabulary where each term corresponds to a set of homogenous visual characteristics (colour and texture features). Since we are going to use a feature space to represent all images, we need a set of visual terms that is able to represent them. Thus, we need to check which visual characteristics are more common in the dataset. For example, if there are a lot of images with a wide range of blue tones we require a larger number of visual terms representing the different blue tones. This draws on the idea that to learn a good high-dimensional visual vocabulary we would benefit from examining the entire dataset to look for the most common set of colour and texture features.

The way we build the high-dimensional visual vocabulary is by clustering the entire dataset and representing each term as a cluster. We follow the approach presented in [14], where the entire dataset is clustered with a hierarchical EM algorithm using a Gaussian mixture model. This approach generates a hierarchy of cluster models that corresponds to a hierarchy of vocabularies with a different number of terms. The ideal number of clusters is selected via the MDL criterion.

4.2 Maximum Entropy Model

Maximum entropy (or logistic regression) is a statistical tool that has been applied to a great variety of fields, e.g. natural language processing, text classification, image annotation. Maximum entropy is used in this paper to model query keywords in the transformed feature space.

We implemented the binomial model, where one class is always modelled relatively to all other classes, and not a multinomial distribution, which would impose a model that does not reflect the reality of the problem: the multinomial model implies that events are exclusive, whereas in our problem keywords are not exclusive. For this reason, the binomial model is a better choice as documents can have more than one keyword assigned.

4.2.1 Over-fitting control: Gaussian Prior

As discussed by Nigan et al [17] and Chen and Rosenfeld [3], maximum entropy models may suffer from overfitting. This is usually because features are high-dimensional and sparse, meaning that the weights can easily push the model density towards particular training data points. Zhang and Oles [24] have also presented a study on the effect of different types of regularisation on logistic regression. As suggested in [17] and [3] we use a Gaussian prior with mean zero and σ^2 variance to prevent the optimisation procedure from overfitting.

4.2.2 Large-Scale Optimization

Newton algorithms need the Hessian matrix to drive the algorithm into a local maximum solution. The computation of the Hessian matrix is very complex because the feature space might have up to ~10,000 dimensions producing the computation of a 10,000×10,000 on each iteration. Thus, algorithms that compute approximations to the Hessian matrix are ideal for the problem at hand. The limited-memory BFGS algorithm proposed by Liu and Nocedal [12] is one of such algorithms. Malouf [16] has compared several optimisation algorithms for maximum entropy and found the limited-memory BFGS algorithm to be the best one. We use the implementation provided by Liu and Nocedal [12].

5. SEMANTIC MULTIMEDIA SIMILARITY

This section describes the semantic space in which images are represented and a similarity metric that relates two documents semantically.

5.1 Semantic Space

In the semantic space multimedia documents are represented as a feature vector of the probabilities of the *T* concepts,

$$\vec{d} = [d_{w_1}, \ldots, d_{w_T}],$$

where each dimension is the probability of concept w_i being present on that document. Note that the vector of concepts is normalised if the similarity metric needs so (normalisation is dependent on the metric). These concepts are extracted by the semantic-multimedia-analyser algorithm described in Section 5.

It is important that the semantic space accommodates as many concepts as possible to be sure that the user idea is represented in that space without losing any concepts. Thus, systems that extract a limited number of concepts are less appropriate. This design requirement pushes us to the research area of metrics on high-dimensional spaces.

We use the tf-idf vector space model. Each document is represented as a vector $\vec{d}$, where each dimension corresponds to the frequency of a given term (concept) w_i from a vocabulary of T terms (concepts). The only difference between our formulation and the traditional vector space model is that we use $P(w_i \mid d)$ instead of the classic term frequency $TF(w_i \mid d)$. This is equivalent because all documents are represented by a high-dimensional vocabulary of length T and

$$P(w_i \mid d) \cdot T \approx TF(w_i \mid d).$$

Thus, to implement a vector space model we set each dimension i of a document vector as

$$d_i = P(w_i \mid d) \cdot IDF(w_i).$$

The inverse document frequency is defined as the logarithm of the inverse of the probability of a concept over the entire collection $\mathcal{D}$,

$$IDF(w_i) = -\log(P(w_i \mid \mathcal{D})).$$

5.2 Cosine Similarity Metric

Documents $\vec{d}$ and queries $\vec{q}$ are represented by vectors of concept probabilities that are computed as was explained before. Several distance metrics exist in the tf-idf representation that compute the similarity between a document $\vec{d_j}$ vector and a query vector $\vec{q}$. We rank documents by their similarity to the query image according to the cosine-distance metric. The cosine similarity metric expression is:

$$\mathrm{sim}\left(\vec{q}, \vec{d}\right) = 1 - \frac{\sum_{i=1}^{T} q_i d_i}{\sqrt{\sum_{i=1}^{T} (q_i)^2} \cdot \sqrt{\sum_{i=1}^{T} (d_i)^2}}.$$

6. EXPERIMENTS

To evaluate our retrieval by semantic example system we tested it on an image dataset and used three performance measures. The following sections will describe the details of the experiments. Note that even though our system supports images and multi-modal data we only present results of preliminary experiments on an image dataset.

6.1 Dataset

This dataset was compiled by Duygulu et al. [4] from a set of COREL Stock Photo CDs. The dataset has some visually similar concepts (*jet*, *plane*, *Boeing*), and some concepts have a limited number examples (10 or less). In their seminal paper, the authors acknowledge that fact and ignored the classes with these problems. The retrieval evaluation scenario consists of a training set of 4500 images and a test set of 500 images. Each image is annotated with 1-5 keywords from a vocabulary of 371 keywords. Only keywords with at least 2 images in the test set were used which reduced the size of the vocabulary to 179 keywords.

6.2 Evaluation

The system was compared to the reference rank constructed from image annotations and random results to provide upper and lower bounds for comparison. Precision and rank correlation measures provide us standard ways to assess our system.

6.2.1 Precision and Average Precision

Precision at 20 (P@20) is the proportion of the first 20 documents that are relevant. The motivation behind this measure is that users will generally only look at the first page of returned results and will rarely scroll beyond 3 pages of results, see [10].

Average precision is the standard measure for comparing a ranked set of results to binary relevance judgments [19]. Conceptually average precision is the area under the precision recall curve. It is calculated by averaging the precision found at every relevant document. The advantage of using average precision as a performance measure is it gives a greater weight to results retrieved early.

These measures are widely used in information retrieval and they consider documents to be relevant or not relevant. However in the current scenario the relevance of a document is difficult to measure. Images can share several concepts or none. The problem is even more complicated because for a particular query an image with one matching concept might be more meaningful than an image with two matching concepts. Nevertheless, we repeated different evaluations where each document was considered correct if it shared one to four concepts with the query. This is limited by the number of manual annotations per image (five) and by the existing images that actually share such a number of concepts.

6.2.2 Rank Correlation

It is difficult to evaluate the results produced by query-by-semantic-example systems: such systems try to match the maximum number of concepts (with associated noise), while humans just match the concepts that are “obvious”. Thus precision may not be a good indication of semantic similarity. To address this issue we created a reference rank that could then be reference correlated to the rank produced by the system. Ideally this reference rank would be constructed manually by multiple users. For the purpose of our evaluation we considered a reference rank constructed with the manual annotations of images and the semantic similarity metric to rank the results. Thus, rank correlation evaluates the semantic analysis according to the similarity metric used to create the reference rank. We use Spearman's rank correlation

$$\rho = 1 - \frac{6\sum_i \Delta_i^2}{n\left(n^2 - 1\right)}, \quad (1)$$

where Δ_i is the difference between the position of the same document d_i on the different ranks, and n is the number of rank positions. ρ quantifies how similar the reference rank is to the rank produced by the system.

6.3 Methodology

In our experiments we trained the 179 concept models on the 4500 training images and used the 500 images for testing. For each testing image we computed the vector of concept's probabilities. Then took each image out and used it as the query example. Then the semantic similarity matched that image with the remaining test images and produces a rank of semantically similar images.

To evaluate our retrieval by semantic example system we used three performance measures: precision at 20, average precision and Spearman's rank correlation [11]. Note that because images have several concepts, images can be semantically matched through multiple concepts. This obviously creates a variety of correct ranks meaning that there is no "correct" unique rank.

7. RESULTS AND DISCUSSION

Query-by-semantic-example is a very special area of multimedia information retrieval that pushes the limits of what can be done with state-of-the-art algorithms. We will now report the most relevant facts that we found in our experiments.

Scalability of the Semantic Query Analysis

The time complexity of the query semantic analysis is a crucial characteristic that we consider to have the same importance as precision. For this reason we carefully chose algorithms that can handle multimedia semantics with little computational complexity. Table 1 illustrates the time required to extract the visual-features and to run the semantic analyses of Section 5. Measures were taken on an AMD Athlon 64 running at 3.7GHz Note that these times are for the inference phase and not for the learning phase.

These times can be further improved because we write the state of all intermediate steps onto disk, which takes much more time than the algorithm itself. On a production system, the data generated from analysing a query example would not have to be written to disk which would greatly improve computational performance. As mentioned before this is an important feature because the system needs to extract the semantics of each example fast and it should also be able to support several users simultaneously.

margHSV 3x3 feature	30 ms
Tamura 3x3 feature	54 ms
Gabor 3x3 feature	378 ms
Semantic annotation (179 concepts)	9 ms

Table 1 – Semantic analysis performance per image.

Rank Correlation

Spearman's rank correlation coefficient (ρ) is a non-parametric measure of correlation. As we compare two ranked lists of results, it is an appropriate measure to quantify how closely the rankings produced by our retrieval-by-semantic-example system correlate to the ground-truth rankings. The mean ρ achieved by our system is 0.27, the rankings produced by our system were significantly different from the ground truth with a confidence of 99.99% (when using the Student's t-distribution approximation), however this is still better than the random mean ρ of 0.25 with a confidence of 99.99%.

Random	0.25
Spearman's RC	0.27

Table 2 – Rank correlation retrieval results.

Retrieval Precision

Retrieval precision evaluation requires a careful analysis of the test data and of the results. Each query image on the test data has five concepts, thus it is very likely that several images on the test set have at least of those five concepts. This means that almost all images are valid for one matching concept. However, when we increase the number of agreeing concepts, the number of relevant images decreases, meaning that it becomes more difficult to retrieve relevant images. This is visible on the random measures that we show on Table 3, where the increase of semantic consistency (matching concepts) the raises the difficult of retrieving relevant images.

Concepts	1	2	3	4
Random MAP	21.02	3.64	1.13	0.36
MAP	23.54	8.16	5.64	2.43
Random Mean P@20	25.09	6.33	1.83	0.52
Mean P@20	31.91	14.29	7.22	2.41

Table 3 – Retrieval precision results (values in percentage).

Looking now at the results we can see that the system performs much better when we need a greater level of semantic consistency. As we increase the number of matching concepts, MAP and Mean P@20 increases more significantly. We believe MAP and P@20 reflect much better results than Spearman's due to the greater weight given to documents returned earlier in the search (in contrast to Spearman's, which gives equal weight to the whole ranking). After a detailed examination of the ground truth we believe this is caused by all annotations having an equal weight. For example: an image of a *bear* in a wood could have annotations, *bear*, *sky* and *tree*. A human assessor may deem *bear* as the most important concept in the image, however, this is not reflected in the annotations.

8. CONCLUSIONS

This paper presented preliminary results concerning query by semantic multimedia example. The conducted experiments allow us to draw interesting conclusions and plan future work.

Query by semantic multimedia example has to work in a high-dimensional semantic space, which is at the same time a strong and a weak characteristic of the system. While the high-dimensionality accommodates a large number of concepts of increasing semantic expressiveness, it also increases the confusion between concepts therefore decreasing accuracy. Different metric measures exist that takes into account the dimensionality of the semantic space but it will always carry a trade-off between accuracy and semantic expressiveness.

The reference rank was produced with the cosine-distance metric between annotated concepts that do not necessarily produce a rank considered correct by a human. We still believe that rank correlation is the best way to access query by semantic example but the reference rank must be constructed by a human.

Figure 2 – A correct but inconsistent result: the query image (left) and the retrieved image (right) share the concept "*water*".

Probably an issue that will never have a solution is related to the fact that semantics are ambiguous by nature. Sometimes, even when a retrieved result is technically correct it is obvious to a human that it should not be at the top of the rank, as the example of Figure 2. A possible way to improve this situation is to have a weighted similarity metric where most-relevant concepts have higher weights.

9. REFERENCES

[1] G. Carneiro, N. Vasconcelos, "Formulating semantic image annotation as a supervised learning problem," IEEE Conf. on Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005.

[2] S.-F. Chang, W. Chen, H. Sundaram, "Semantic visual templates: linking visual features to semantics," ICIP, Chicago, IL, USA, 1998.

[3] S. F. Chen, R. Rosenfeld, "A Gaussian prior for smoothing maximum entropy models," Carnegie Mellon University, Pittsburg, PA February 1999.

[4] P. Duygulu, K. Barnard, N. de Freitas, D. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," European Conf. on Computer Vision, Copenhagen, Denmark, 2002.

[5] S. L. Feng, V. Lavrenko, R. Manmatha, "Multiple Bernoulli relevance models for image and video annotation," IEEE Conf. on Computer Vision and Pattern Recognition, Cambridge, UK, 2004.

[6] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, P. Yanker, "Query by image and video content: the QBIC system," *IEEE Computer*, vol. 28, pp. 23-32, 1995.

[7] X. He, O. King, W.-Y. Ma, M. Li, H.-J. Zhang, "Learning a semantic space from user's relevance feedback for image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 39- 48, 2003.

[8] D. Heesch, S. Rüger, "Interaction models and relevance feedback in content-based image retrieval," in *Semantic-Based Visual Information Retrieval*, Y.-J. Zhang, Ed.: Idea-Group, 2006.

[9] D. Hull, "Using statistical testing in the evaluation of retrieval experiments," ACM SIGIR Conference, Pittsburgh, PA, USA, 1993.

[10] iProspect, "Search Engine User Behavior Study," April 2006.

[11] E. L. Lehmann, *Nonparametrics: statistical Methods Based on Ranks* San Francisco: Holden-Day, 1975.

[12] D. C. Liu, J. Nocedal, "On the limited memory method for large scale optimization," *Mathematical Programming B*, vol. 45, pp. 503-528, 1989.

[13] Y. Lu, C. Hu, X. Zhu, H. Zhang, Q. Yang, "A unified framework for semantics and feature based relevance feedback in image retrieval systems," ACM Multimedia, Los Angeles, CA, USA, 2000.

[14] J. Magalhães, S. Rüger, "Logistic regression of generic codebooks for semantic image retrieval," Int'l Conf. on Image and Video Retrieval, Phoenix, AZ, USA, 2006.

[15] J. Magalhães, S. Rüger, "Information-theoretic semantic multimedia indexing," in *ACM Conference on Image and Video Retrieval*. Amsterdam, Holland, 2007.

[16] R. Malouf, "A comparison of algorithms for maximum entropy parameter estimation," Sixth Conf. on Natural Language Learning, Taipei, Taiwan, 2002.

[17] K. Nigam, J. Lafferty, A. McCallum, "Using Maximum Entropy for Text Classification," IJCAI - Workshop on Machine Learning for Information Filtering, Stockholm, Sweden, 1999.

[18] N. Rasiwasia, N. Vasconcelos, P. Moreno, "Query by Semantic Example," CIVR, Phoenix, AZ, USA, 2006.

[19] E. Voorhees, D. Harman, "Overview of the 8th Text REtrieval Conference TREC-8," TREC-8, Gaithersburg, MD, 1999.

[20] J. Z. Wang, J. Li, G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 947-963, 2001.

[21] C. Yang, M. Dong, F. Fotouhi, "Semantic feedback for interactive image retrieval," Int'l Multimedia Modelling Conference, Singapore, 2005.

[22] A. Yavlinsky, E. Schofield, S. Rüger, "Automated image annotation using global features and robust nonparametric density estimation," Int'l Conf. on Image and Video Retrieval, Singapore, 2005.

[23] C. Zhang, T. Chen, "An active learning framework for content-based information retrieval," *IEEE Transactions on Multimedia*, vol. 4, pp. 260- 268, 2002.

[24] T. Zhang, F. J. Oles, "Text Categorization Based on Regularized Linear Classification Methods," *Information Retrieval*, pp. 5-31, 2001.

[25] X. S. Zhou, T. S. Huang, "Unifying keywords and visual contents in image retrieval," *IEEE Multimedia*, vol. 9, pp. 23-33, 2002.