

QUERY SHIFTING BASED ON BAYESIAN DECISION THEORY FOR CONTENT-BASED IMAGE RETRIEVAL

Giorgio Giacinto and Fabio Roli

Dept. of Electrical and Electronic Engineering - University of Cagliari
Piazza D'Armi 09123 Cagliari, Italy
Tel: +39 070 675 5752 Fax: +39 070 675 5782
e-mail {giacinto,roli}@diee.unica.it

Abstract. *Despite the efforts to reduce the so-called semantic gap between the user's perception of image similarity and feature-based representation of images, the interaction with the user remains fundamental to improve performances of content-based image retrieval systems. To this end, relevance feedback mechanisms are adopted to refine image-based queries by asking users to mark the set of images retrieved in a neighbourhood of the query as being relevant or not. In this paper, Bayesian decision theory is used to compute a new query whose neighbourhood is more likely to fall in a region of the feature space containing relevant images. The proposed query shifting method outperforms two relevance feedback mechanisms described in the literature. Reported experiments also show that retrieval performances are less sensitive to the choice of a particular similarity metric when relevance feedback is used.*

1. Introduction

The availability of large image and video archives for many applications (art galleries, picture and photograph archives, medical and geographical databases, etc.) demands advanced query mechanisms that address perceptual aspects of visual information. To this end, a number of image retrieval techniques based on image content, where the visual content of images is captured by extracting low-level features based on color, texture, shape, etc., have been developed [4],[19]. Content-based queries are often expressed by visual examples in order to retrieve from the database all images that are “similar” to the examples. The retrieval process is usually performed by a k - nn search in the feature space using the Euclidean metric [4].

It is easy to see that the effectiveness of a content-based image retrieval system (CBIR) strongly depends on the choice of the set of visual features and on the choice of the “metric” used to model the user's perception of image similarity. The gap between user's perception of image similarity and feature-based image representation is usually small for databases related to tasks where the semantic description of the images is reasonably well defined. For example, data bases of lithographs, frontal views of faces, outdoor pictures, etc. [19]. For this kind of databases, a pair of images

that the user judges as being similar to each other is often represented by two near points in the feature space.

However, no matter how suitable for the task at hand the features and the similarity metric have been designed, the set of retrieved images often fits the user's needs only partly. Typically, different users may categorise images according to different semantic criteria [1]. Thus, if we allow different users to mark the images retrieved with a given query as "relevant" or "non-relevant", different subsets of images will be marked as "relevant", and the intersection of such subsets is usually non-empty. Accordingly, the need for mechanisms to "adapt" the CBIR system response based on some "feedback" from the user is widely recognised.

A number of techniques aimed at exploiting such relevance feedback have been proposed in the literature [2],[3],[6],[7],[9],[10],[11],[12],[13],[14],[16],[18]. As discussed in Section 2, they are based on the fact that the user does not know the actual distribution of images in the feature space, nor the feature space itself, nor the similarity metric employed.

In this paper, Bayesian decision theory is used to compute a new query point based on relevance feedback from the user. The basic idea behind our proposal is the local estimation of the decision boundary between the "relevant" and "non relevant" regions of the neighbourhood of the original query. The new query is then placed at a suitable distance from such boundary, on the side of the sub-region containing relevant images. A similar query shifting mechanism was proposed by the authors in [7], where the query shifting computation was derived by heuristics. In this paper, the computation of the new query is placed in the framework of Bayesian decision theory.

In section 2, a brief overview of relevance feedback techniques for CBIR is given. The proposed relevance feedback method is described in Section 3. Experiments with two image datasets are reported in Section 4. The reported results show that the proposed method outperforms two relevance feedback mechanisms recently described in the literature. Section 4 also points out that, when relevance feedback is performed, retrieval performances are less sensitive to the choice of a particular similarity metric.

2. Relevance Feedback for CBIR

It is well known that information retrieval system performances can be improved by user interaction mechanisms. This issue has been studied thoroughly in the text retrieval field, where the relevance feedback concept has been introduced [15]. Techniques developed for text retrieval systems should be suitably adapted to content based image retrieval, on account of differences in both feature number and meaning, and in similarity measures [10],[13].

Basically, relevance feedback strategies are motivated by the observation that the user is unaware of the distribution of images in the feature space, nor of the feature space itself, nor of the similarity metric. Therefore, relevance feedback techniques proposed in the literature involve the optimisation of one or more CBIR components, e.g., the formulation of a new query, the modification of the similarity metric, or the transformation of the feature space. Query reformulation is motivated by the observation that the image used to query the database may be placed in a region of the

feature space that is "far" from the one containing images that are relevant to the user. A query shifting technique for CBIR based on the well known Rocchio formula developed in the text retrieval field [15] has been proposed in [13]. The estimation of probability densities of individual features for relevant and non relevant images is used in [11] to compute a new query. The new query is determined by randomly drawing individual feature components according to the estimated distributions.

In order to optimise the similarity metric to user interests, many CBIR systems rely on parametric similarity metrics, whose parameters are optimised by relevance feedback. Theoretical frameworks involving both the computation of a new query and the optimisation of the parameters of similarity metric have been proposed in the literature [9],[14]. A linear combination of different similarity metrics, each suited for a particular feature set, has been proposed in [18]. Relevance feedback information is then used to modify the weights of the combination to reflect different feature relevance. Santini and Jain also proposed a parametrized similarity measure updated according to feedback from the user [16]. Rather than modifying the similarity metric, Frederix et al. proposed a transformation of the feature space by a logistic regression model so that relevant images represented in the new feature space exhibit higher similarity values [6]. A probabilistic feature relevance scheme has been proposed in [12], where a weighted Euclidean distance is used.

A different perspective has been followed in [3] where relevance feedback technique based on the Bayesian decision theory was first proposed. The probability of all images in the database of being relevant is estimated, and images are presented to the user according to the estimated probability.

3. Query shifting by Bayesian decision theory

3.1 Problem formulation

Let us assume first that the database at hand is made up of images whose semantic description is reasonably well defined. In these cases it is possible to extract a set of low level features, such that a pair of images judged by the user as being similar to each other is represented by two near points in the feature space. Let us also assume that the user wishes to retrieve images belonging to a specific *class*, that is, she/he is interested in performing a so-called "category" search [19]. As different users have different perceptions of similarity depending on the goal they are pursuing, for a given query, different users may identify different subsets of relevant images. According to the first hypothesis, each subset of relevant images identifies a region in the feature space. Relevance feedback is thus needed to locate the region containing relevant images for a given user.

The user marks the images retrieved by the k-nn search as being relevant or not, so that the neighbourhood of the query in the feature space is subdivided into a relevant and a non-relevant region. Our approach is based on the local estimation of the boundary between relevant and non-relevant images belonging to the neighbourhood.

Then a new query is computed so that its neighbourhood is more likely to be contained in the relevant region.

In order to illustrate our approach, let us refer to the example shown in Figure 1. The boundary of the region that contains the relevant images that the user wishes to retrieve is depicted in the figure. It is worth noting that this boundary is not known a-priori because its knowledge would require the user to mark all images contained in the database. \mathbf{Q}_0 is the initial query provided by the user to perform the k -nn search. The neighbourhood $N(\mathbf{Q}_0)$ of \mathbf{Q}_0 does not fall entirely inside the region of relevant images because it contains both relevant and non-relevant images.

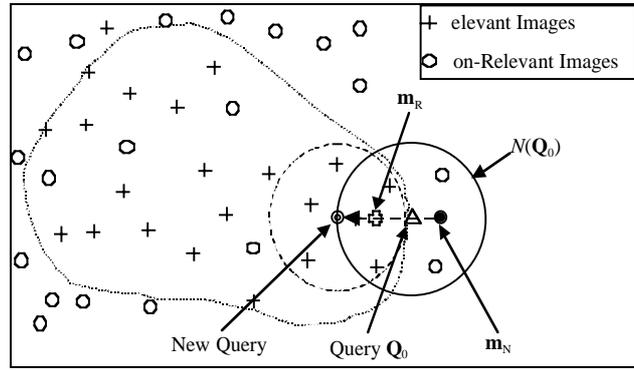


Fig. 1. The proposed query shifting method: an example. The boundary of the region containing the images that are relevant to user query \mathbf{Q}_0 is depicted by the dotted line. The initial query \mathbf{Q}_0 and the neighbourhood $N(\mathbf{Q}_0)$ related to the k -nn search ($k = 5$) are depicted. A new query computed in the $\mathbf{m}_R - \mathbf{m}_N$ direction is shown such that its neighbourhood (dashed line) is contained in the relevant region. \mathbf{m}_R and \mathbf{m}_N are the mean vectors of relevant and non-relevant image subsets, respectively, retrieved with the initial query \mathbf{Q}_0 .

The decision boundary between relevant and non relevant images belonging to $N(\mathbf{Q}_0)$ can be estimated as explained in the following.

Let \mathbf{I} be a feature vector representing an image in a d -dimensional feature space. Let $I_R(\mathbf{Q}_0)$ and $I_N(\mathbf{Q}_0)$ be the sets of relevant and non-relevant images, respectively, contained in $N(\mathbf{Q}_0)$. The mean vectors of relevant and non-relevant image, \mathbf{m}_R and \mathbf{m}_N , can be computed as follows

$$\mathbf{m}_R = \frac{1}{k_R} \sum_{\mathbf{I} \in I_R(\mathbf{Q}_0)} \mathbf{I}, \quad \mathbf{m}_N = \frac{1}{k_N} \sum_{\mathbf{I} \in I_N(\mathbf{Q}_0)} \mathbf{I} \quad (1)$$

where k_R and k_N are the sizes of relevant and non-relevant image sets, respectively ($k_R + k_N = k$). The average variance of relevant and non relevant images can be computed as follows:

$$\sigma^2 = \frac{1}{k} \left(\sum_{\mathbf{I} \in I_R(\mathbf{Q}_0)} (\mathbf{I} - \mathbf{m}_R)^T (\mathbf{I} - \mathbf{m}_R) + \sum_{\mathbf{I} \in I_N(\mathbf{Q}_0)} (\mathbf{I} - \mathbf{m}_N)^T (\mathbf{I} - \mathbf{m}_N) \right) \quad (2)$$

Let us assume that relevant and non-relevant images in $N(\mathbf{Q}_0)$ are normally distributed with means \mathbf{m}_R and \mathbf{m}_N and equal variance σ^2 . Then, according to the

Bayesian decision theory, the decision surface between these two “classes” of images is orthogonal to the line linking the means and passes through a point \mathbf{x}_0 defined by the following equation [5]:

$$\mathbf{x}_0 = \frac{1}{2}(\mathbf{m}_R + \mathbf{m}_N) - \frac{\sigma^2}{\|\mathbf{m}_R - \mathbf{m}_N\|^2} \ln \frac{P(\omega_R)}{P(\omega_N)} (\mathbf{m}_R - \mathbf{m}_N) \quad (3)$$

where the priors $P(\omega_R)$ and $P(\omega_N)$ are related to the images belonging to $N(\mathbf{Q}_0)$ and can be estimated as the fraction of relevant and non relevant images in $N(\mathbf{Q}_0)$, respectively.

When the prior probabilities are equal, i.e., half of the images of $N(\mathbf{Q}_0)$ are relevant, \mathbf{x}_0 is halfway between the means, while it moves away from the more likely mean in the case of different priors. In \mathbf{x}_0 the posterior probabilities for the two classes are equal, while points with higher values of posterior probability for class ω_R are found by moving away from \mathbf{x}_0 in the $(\mathbf{m}_R - \mathbf{m}_N)$ direction (if we move in the opposite direction, higher posteriors for ω_N are obtained). Therefore, as clearly shown in Figure 1, candidate query points that could improve retrieval performances, are those located on the line connecting \mathbf{m}_R and \mathbf{m}_N . In particular the new query point should be selected in the $\mathbf{m}_R - \mathbf{m}_N$ direction so that its neighbourhood is contained in the relevant region.

3.2 Query Shifting Computation

The rationale behind the query computation proposed hereafter can be briefly explained as follows. The desired result is to have the neighbourhood of the new query totally contained in the relevant region of the feature space. Therefore we first hypothesise an optimal location for the desired neighbourhood and then we will compute the query that can be associated with such a neighbourhood.

An optimal neighbourhood can be obtained by shifting the neighbourhood of \mathbf{Q}_0 in the $\mathbf{m}_R - \mathbf{m}_N$ direction until it contains only relevant images (see Figure 1). Let $\mathbf{m}_R^{(1)}$ be the mean vector of relevant images captured by this shifted neighbourhood. We propose to use this point as the new query.

Exploiting the hypotheses of section 3.1, and following the above rationale, let us derive formally the computation of the new query. Let us define the shifted neighbourhood as the neighbourhood whose images satisfy the following properties: i) the mean vectors of relevant and non-relevant images are collinear with the mean vectors of $N(\mathbf{Q}_0)$ and their distance is constant, i.e.

$$\mathbf{m}_R^{(1)} - \mathbf{m}_N^{(1)} = \mathbf{m}_R^{(0)} - \mathbf{m}_N^{(0)} \quad (4)$$

where superscripts (0) and (1) refer to original neighbourhood position and the shifted one, respectively; ii) the average variance of relevant and non-relevant images is always equal to σ^2 ; iii) the boundary between relevant and non relevant images estimated according to Equation (3) for any shifted neighbourhood, coincides with the boundary computed using the original neighbourhood, i.e., \mathbf{x}_0 represents a point of the actual boundary between relevant and non relevant images. Accordingly, the location

of the point \mathbf{x}_0 (Equation 3) can be computed using either neighbourhood (0) or (1). By making equal the two computations of \mathbf{x}_0 , the following relation holds:

$$\frac{1}{2}(\mathbf{m}_R^{(1)} + \mathbf{m}_N^{(1)} - \mathbf{m}_R^{(0)} - \mathbf{m}_N^{(0)}) = \frac{\sigma^2}{\|\mathbf{m}_R^{(0)} - \mathbf{m}_N^{(0)}\|^2} \left(\ln \frac{P^{(1)}(\omega_R)}{P^{(1)}(\omega_N)} - \ln \frac{P^{(0)}(\omega_R)}{P^{(0)}(\omega_N)} \right) (\mathbf{m}_R^{(0)} - \mathbf{m}_N^{(0)}) \quad (5)$$

It is worth noting that the two neighbourhoods capture different fractions of relevant and non relevant images, i.e., in the above formula the priors are different.

To simplify the above expression and avoid infinite results, let us substitute the logs with the following first-order approximation:

$$\ln \frac{P(\omega_R)}{P(\omega_N)} \cong \frac{k_R - k_N}{\max(k_R, k_N)} \quad (6)$$

where each prior is estimated as the fraction of relevant and non relevant images contained in the neighbourhood ($P(\omega_R) + P(\omega_N) = 1$).

If we let the superscript (0) indicate the data computed from the relevance information provided by the user, Equation 5 let us compute the position of the mean $\mathbf{m}_R^{(1)}$ for which the related neighbourhood does not contain any non-relevant image, i.e. $P^{(1)}(\omega_N)$ is equal to 0. We shall select this point as the new query. By substituting Equation 6 in Equation 5 and expressing $\mathbf{m}_N^{(1)}$ as a function of $\mathbf{m}_R^{(1)}$, $\mathbf{m}_R^{(0)}$ and $\mathbf{m}_N^{(0)}$ according to Equation 4, and letting $P^{(1)}(\omega_N) = 0$, the point $\mathbf{m}_R^{(1)}$ where the new query \mathbf{Q}_I should be placed can be computed as follows

$$\mathbf{m}_R^{(1)} = \mathbf{m}_R^{(0)} + \frac{\sigma^2}{\|\mathbf{m}_R^{(0)} - \mathbf{m}_N^{(0)}\|^2} \left(1 - \frac{k_R^{(0)} - k_N^{(0)}}{\max(k_R^{(0)}, k_N^{(0)})} \right) (\mathbf{m}_R^{(0)} - \mathbf{m}_N^{(0)}) \quad (7)$$

Summing up, the query computed by Equation (7) coincides with $\mathbf{m}_R^{(0)}$ only when all images in $N(\mathbf{Q}_0)$ are relevant. Otherwise the larger the fraction of non-relevant images in the neighbourhood of the original query, the further the new query from the original neighbourhood.

It is worth to point out the main difference between the proposed query computation and other methods proposed in the literature. Usually the new query is computed as the solution of the minimisation of the average distance of the query with all the retrieved images, where relevant images have larger weights than non-relevant images. The new query is thus the weighted average of all retrieved images, the weights being related to the degree of relevance [14]. Therefore the new query is the "optimum" query with respect to the retrieved images. On the other hand the proposed mechanism is based on a local model of the distribution of relevant and non relevant images. This model is used to "optimise" the location of the neighbourhood of the new query with respect to the local boundary between relevant and non relevant images. To this end the new query is computed at a distance from the boundary

proportional to the neighbourhood "size" expressed in terms of the ratio between the variance and the distance between means.

Further discussions on the validity of the proposed approach in comparison with other approaches proposed in the literature, are out of the scope of the present paper.

4. Experimental Results

In order to test the proposed method and compare it with other methods described in the literature, two image databases have been used: the MIT database and a database contained in the UCI repository.

The MIT database was collected by the MIT Media Lab (<ftp://whitechapel.media.mit.edu/pub/VisTex>). This database contains 40 texture images that have been processed as described in [13]. Images have been manually classified into fifteen classes. Each of these images has been subdivided into sixteen non-overlapping images, obtaining a data set with 640 images. Sixteen Gabor filters were used to characterise these images, so that each image is represented by a 16-dimensional feature vector .

The database extracted from the UCI repository (<http://www.cs.uci.edu/mlearn/MLRepository.html>) consists of 2,310 outdoor images. The images are subdivided into seven data classes (brickface, sky, foliage, cement, window, path, and grass). Nineteen colour and spatial features characterise each image. (Details are reported on the UCI web site).

For each dataset, a normalisation procedure has been performed, so that each feature takes values in the range between 0 and 1. This normalisation procedure is necessary when the Euclidean distance metric is used.

For both databases, each image is used as a query and the top twenty nearest neighbours are returned. Relevance feedback is performed by marking images belonging to the same class of the query as relevant, and all other images in the top twenty as non-relevant. This experimental set up affords an objective comparison among different methods and is currently used by many researchers [11], [12],[13].

Tables 1 and 2 report the results of the proposed method on the two selected datasets in terms of average percentage retrieval precision and Average Performance Improvement (API). Precision is measured as the ratio between the number of relevant retrievals and the number of total retrievals averaged over all the queries. API is computed averaging the following ratio over all the queries:

$$\frac{\text{relevant retrievals}(n+1) - \text{relevant retrievals}(n)}{\text{relevant retrievals}(n)}$$

where $n = 0, 1, \dots$ is the number of feedbacks performed. In the reported experiments, n equals 1, because the relative performances of the compared methods does not change significantly by increasing the number of feedbacks performed.

For the sake of comparison, retrieval performances obtained with other methods recently described in the literature are also reported, namely the RFM (Relevance Feedback Method) [13] and the PFRL (Probabilistic Feature Relevance Learning) [12]. PFRL is a probabilistic feature relevance feedback method aimed at weighting each feature according to the information extracted from the relevant images. This

method uses the Euclidean metric to measure the similarity between images. RFM is an implementation of the Rocchio formula for CBIR, that is, it implements the query shifting strategy. It is worth noting that RFM uses the cosine metric to compute similarity between images. Therefore, a different normalisation procedure is performed on the data sets in order to adapt features to the cosine metric.

The first columns of Tables 1 and 2 report the average percentage retrieval precision without feedback step. It is worth noting that the reported differences in performances depend on the different similarity metrics used. These results show that the cosine metric is more suited than the Euclidean metric to the MIT data set, while the reverse is true for the UCI data set. This points out that, if no relevance feedback mechanism is used, retrieval performances are highly sensitive to the selected similarity metric.

Table 1. Retrieval Performances for the MIT data set. Average percentage retrieval precision and Average Performance Improvement (API) are reported.

RF mechanism	1st retrieval	2nd retrieval with RF	API
Rocchio	83.74%	90.23%	13.53
PFRL	79.24%	85.48%	12.70
Bayesian query shifting	79.24%	91.11%	28.79

Table 2. Retrieval Performances for the UCI data set. Average percentage retrieval precision and Average Performance Improvement (API) are reported.

RF mechanism	1st retrieval	2nd retrieval with RF	API
Rocchio	86.39%	91.95%	15.33
PFRL	90.21%	94.56%	7.66
Bayesian query shifting	90.21%	96.24%	15.64

The second columns of Tables 1 and 2 report the average percentage retrieval precision after relevance feedback. The proposed relevance feedback method always outperformed the PFRL and the Rocchio formula. It is worth noting that while the Rocchio formula and the PFRL relies on some parameters that must be chosen by heuristics, the proposed method is based only on statistical estimates in the neighbourhood of the original query. However, the limited experimentation carried out does not allow to draw definitive conclusions.

A comparison between the PFRL and the proposed query shifting method shows that query shifting is more suited to relevance feedback than feature weighting alone. This is also confirmed by the results reported in [8], where PFRL performances are improved by combining PFRL with a query shifting mechanism.

With regard to the results on the MIT data set, it should be noted that although the method based on the Rocchio formula obtained a larger number of relevant images in the first retrieval the proposed query shifting method outperformed it when relevance feedback was used. Therefore, one can argue that retrieval performances provided by the proposed relevance feedback method are less sensitive to the choice of the similarity metric.

The above conclusions are also confirmed when comparing the average performance improvements (API). Our method provided the largest performance

improvements on both data sets. In particular, the advantages of the proposed method are more evident on the MIT data set.

References

1. Bhanu, B., Dong, D.: Concepts Learning with Fuzzy Clustering and Relevance Feedback. In: Petra, P. (Ed.): Machine Learning and Data Mining in Pattern Recognition. LNAI 2123, Springer-Verlag, Berlin (2001) 102-116
2. Ciocca G, Schettini R.: Content-based similarity retrieval of trademarks using relevance feedback. Pattern Recognition, **34**(8) (2001) 1639-1655
3. Cox, I.J., Miller, M.L., Minka T.P., Papatomas T.V., Yianilos, P.N.: The Bayesian image retrieval system, *PicHunter*: theory, implementation, and psychophysical experiments. IEEE Trans. on Image Processing **9**(1) (2000) 20-37
4. Del Bimbo A.: Visual Information Retrieval. Morgan Kaufmann Pub. Inc., San Francisco, CA (1999)
5. Duda R.O., Hart P.E., Stork D.G.: Pattern Classification. J. Wiley & Sons (2000)
6. Frederix G., Caenen G., Pauwels E.J.: PARISS: Panoramic, Adaptive and Reconfigurable Interface for Similarity Search. Proc. of ICIP 2000 Intern. Conf. on Image Processing. WA 07.04, vol. III (2000) 222-225
7. Giacinto, G., Roli, F., Fumera, G.: Content-Based Image Retrieval with Adaptive Query Shifting. In: Petra, P. (Ed.): Machine Learning and Data Mining in Pattern Recognition. LNAI 2123, Springer-Verlag, Berlin, (2001) 337-346
8. Hesterkamp DR, Peng J, Dai HK.: Feature relevance learning with query shifting for content-based image retrieval. In Proc. of the 15th IEEE International Conference on Pattern Recognition (ICPR 2000), vol 4. IEEE Computer Society (2000) 250-253
9. Ishikawa Y., Subramanys R., Faloutsos C.: MindReader: Querying databases through multiple examples. In Proceedings. of the 24th VLDB Conference (1998) 433-438
10. McG Squire D, Müller W., Müller H., Pun T.: Content-based query of image databases: inspirations from text retrieval. Pattern Recognition Letters **21**(13-14) (2000) 1193-1198
11. Nastar C., Mitschke M., Meilhac C.: Efficient query refinement for Image Retrieval. Proc. of IEEE Conf. Computer Vision and Pattern Recognition, CA (1998) 547-552
12. Peng J., Bhanu B., Qing S.: Probabilistic feature relevance learning for content-based image retrieval. Computer Vision and Image Understanding **75**(1-2) (1999) 150-164
13. Rui Y., Huang T.S., Mehrotra S.: Content-based image retrieval with relevance feedback: in MARS. In Proceedings of the IEEE International Conference on Image Processing, IEEE Press (1997) 815-818
14. Rui Y., Huang T.S.: Relevance Feedback Techniques in Image retrieval. In Lew M.S. (ed.): Principles of Visual Information Retrieval. Springer-Verlag, London, (2001) 219-258
15. Salton G., McGill M.J.: Introduction to modern information retrieval. McGraw-Hill, New York (1988)
16. Santini S., Jain R.: Integrated browsing and querying for image databases. IEEE Multimedia **7**(3) (2000) 26-39
17. Santini S., Jain R.: Similarity Measures. IEEE Trans. on Pattern Analysis and Machine Intelligence **21**(9) (1999) 871-883
18. Sclaroff S., La Cascia M., Sethi S., Taycher L.: Mix and Match Features in the ImageRover search engine. In Lew M.S. (ed.): Principles of Visual Information Retrieval. Springer-Verlag, London (2001) 219-258
19. Smeulders A.W.M., Worring M., Santini S., Gupta A., Jain R.: Content-based image retrieval at the end of the early years. IEEE Trans. on Pattern Analysis and Machine Intelligence **22**(12) (2000) 1349-1380