14

Instance-Based Relevance Feedback in Image Retrieval Using Dissimilarity Spaces

G. Giacinto and F. Roli

Department of Electrical and Electronic Engineering – University of Cagliari, Italy

Summary. High-retrieval precision in content-based image retrieval can be attained by adopting relevance feedback mechanisms. The user marks all the retrieved images as being either relevant or not, then the search engine exploits this relevance feedback to adapt the search to better meet user's needs. The main difficulties in exploiting relevance information are (a) the gap between user perception of similarity and the similarity computed in the feature space used for the representation of image content and (b) the availability of few training data (users typically label a few dozen of images. At present, SVM are extensively used to learn from relevance feedback due to their capability of effectively tackling the above difficulties. As the performances of SVM depend on the tuning of a number of parameters, in this chapter a different approach to relevance feedback is proposed. First, images are represented in the dissimilarity space made up of the dissimilarities from the set of relevant images. Then a relevance score is computed in terms of the distance from the nearest nonrelevant image, and the distance from the nearest relevant one. Images are ranked according to this score and the top k images are displayed. Reported results show that the performances of the proposed approach are comparable to the highest performances that can be attained by SVM by suitably tuning the learning parameters.

14.1 Introduction

Search engines are becoming increasingly popular as far as the amount of information available in digital form makes it nearly impossible to users to browse inside predefined categories, as each category may include thousands of items. On the other hand, search engines employ techniques from the artificial intelligence domain that allows retrieving data not only according to the associated metadata used to describe them, but also in term of their content [16]. To this end, a suitable model of the content is needed. While search engines for textual documents have reached a good level of maturity (e.g., the search engine employed by Google), search engines for images are still far to be accepted by the average user. The main reason is that images

G. Giacinto and F. Roli: *Instance-Based Relevance Feedback in Image Retrieval Using Dissimilarity Spaces*, Studies in Computational Intelligence (SCI) **73**, 419–436 (2008)
www.springerlink.com

convey a vast amount of information (images have been used as the first mean of communication among humans) [33]. Thus every description of an image is inherently subjective and partial, and search for images based on keywords may fits users' needs only partially. This fact explains why researchers are investigating techniques to retrieve images from databases where the search is based on image content. This research topic is usually referred to as content-based image retrieval (CBIR) and it is currently attracting many researchers from different fields [16, 33].

As it is very difficult to capture the complex semantics of an image, the vast majority of CBIR techniques relies on the representation of images by low-level features, e.g., color, texture, shape, etc., [8,16,33]. As a consequence, content-based queries are typically expressed by visual examples in order to retrieve from the database all images that are "similar" to the examples. It is easy to see that the effectiveness of CBIR techniques 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. A number of metrics have been proposed in the literature to adequately measure (dis)similarities in a given feature space [16, 30, 33].

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. It is easy to see that different users may categorize images according to different semantic criteria [3, 41]. Thus, if we allow different users to mark the images retrieved with a given query as relevant or nonrelevant, different subsets of images will be marked as relevant. Accordingly, the need for mechanisms to adapt the CBIR system response based on some feedback from the user is widely recognized [16, 33, 41].

This issue has been studied thoroughly in the text retrieval field, where the relevance feedback concept has been introduced [29]. However, as far as effective document models have been devised, relevance feedback became less attractive so that they are rarely used in the text retrieval field. On the other hand, relevance feedback has been recognized to be necessary in CBIR due to the inherent subjectivity in measuring the degree of relevance of an image with respect to a given query.

A number of relevance feedback techniques have been proposed in the literature to date [41]. Early works on relevance feedback have been formulated in terms of the optimization of one or more CBIR components, e.g., the formulation of a new query and/or the modification of the similarity metric to take into account the relevance of each feature to the user query [15, 25, 28]. Other CBIR systems employ parametric similarity metrics whose parameters are computed from relevance feedback [31, 32].

More recently, relevance feedback has been formulated in terms of a classification problem [37–40]. This formulation requires a careful design, as the number of training samples is typically small (the user is asked to mark as being relevant or not a number of images in the order of few dozens), while the number of features used to represent image content can be large. In addition, the problem can be either formulated as a two-class problem (relevant vs. nonrelevant), or as a $(1 + x)$-class problem. This second formulation, a.k.a. "biased learning," takes into account that the total number of classes in the image database in unknown, but the user is only interested in one class [40]. Two learning techniques are widely used in this context: support vector machines (SVM), and discriminant analysis (DA).

SVM are quite popular in the pattern recognition field as they can handle the above two issues (i.e., small training sets and high dimensionality), and the reported results are quite good if compared to other techniques. However, it should be noted that the choice of the most appropriate SVM parameters for the problem at hand is far from being trivial. Different values of the learning parameters as well as different choices of the kernel may lead to different performances. This aspect is almost neglected in research papers, but it needs to be tackled in order to use SVM in an operational environment.

Discriminant analysis on the other hand has been formulated in a "biased" environment in account of the lack of knowledge on the number of classes. Reported results are quite good even if, as for SVM, a number of learning parameters must be set appropriately in order to get valuable results.

In this chapter, we present a relevance feedback mechanism based on the representation of the images in the database in terms of their (dis)similarities from a *representation set*. The use of the dissimilarity representation of objects has been recently studied in the pattern recognition field where it provided alternative solutions to a number of pattern recognition problems [10, 22, 23]. In particular, it is suited in cases when it is difficult to provide a good feature representation of data, while it can be easier to provide a set of dissimilarities from a set of representative objects. It has also been shown that the dissimilarity representation of data can be viewed as an application of case-based reasoning [26].

As the low-level feature representation of images does not always allow for capturing the user concept of similarity, the dissimilarity representation may represent an approach to exploit relevance feedback [5, 12, 20]. It has been shown that the dissimilarity representation of images allows bridging the gap between low-level (feature) representation of images and the user's perception of similarity, as low-level features are used an intermediate step between images and relevance feedback computation.

In order to exploit the benefit of the dissimilarity representation of images, we compute a relevance score for each image of the database in terms of the kth distance from the set of relevant images, and the kth distance from the set of nonrelevant images retrieved so far. For this reason we called this approach "instance based" as, for each image of the database, its relevance score depends on the distance from one relevant image, and the distance from one nonrelevant image.

This chapter is organized as follows. In Sect. 14.2, a brief review of the related works on relevance feedback is presented. In Sect. 14.3, the dissimilarity representation of images in the context of CBIR systems is proposed.

The proposed relevance feedback mechanism based on the nearest neighbor paradigm is described in Sect. 14.4. Experimental results on an image data set are reported in Sect. 14.5. Reported results show that the performances of the proposed method can be compared to other relevance feedback mechanisms described in the literature. Conclusions are drawn in Sect. 14.6.

14.2 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 [29]. Techniques developed for text retrieval should be suitably adapted to CBIR, on account of differences in both feature number and meaning, and in similarity measures [17, 27, 28].

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 optimization of one or more CBIR components, e.g., the formulation of a new query and/or the modification of the similarity metric to take into account the relevance of each feature to the user query.

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 has been proposed in [27].

Relevance feedback is used in many CBIR systems to optimize a parametric similarity metric. A linear combination of different similarity metrics, each suited for a particular feature set, has been proposed in [32]. Relevance feedback information is then used to modify the weights of the combination to reflect different feature relevance. Santini and Jain also proposed a parameterized similarity measure updated according to feedback from the user [31]. 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 [11]. A probabilistic feature relevance scheme has been proposed in [25], where a weighted Euclidean distance is used. Theoretical frameworks involving both the computation of a new query and the optimization of the parameters of similarity metric have been also proposed [15, 28].

A different perspective has been followed in [6] 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. Bayesian decision theory also inspired a query shifting approach aimed at computing a new query whose k-nearest neighbors belongs to the *relevant* region of the feature space [13].

More recently, relevance feedback has been formulated in terms of a classification problem [37–40]. The problem has been either formulated as a two-class problem (relevant vs. nonrelevant), or as a $(1 + x)$-class problem. This second formulation, a.k.a. "biased learning," takes into account that the total number of classes in the image database in unknown, but the user is only interested in one class [40]. Two learning techniques are widely used in this context: SVM, and discriminant analysis (DA). Reported results showed that these approaches allow for attaining better results than those provided by early relevance feedback approaches. A number of papers proposed different approaches based on the above paradigms. Advanced techniques as well as an extended overview of the two techniques can be found in [35] and [34].

The use of the dissimilarity representation for relevance feedback in CBIR has been first proposed in [12]. Other researchers independently proposed different approaches based on the dissimilarity representation of data where SVM are used as the final classification tool [5, 20]. Reported results showed that this representation of images allows bridging the gap between low-level (feature) representation of images and the user's perception of similarity.

14.3 Dissimilarity Representation of Images

Dissimilarity representation of data has been proposed in the pattern recognition field as an alternative approach in representing data w.r.t. the feature representation of data [23]. Instead of representing patterns in terms of a feature vector, patterns are represented by a vector of (dis)similarities from a set of *representative* data. This approach is well suited for those applications where it is difficult to provide a suitable representation of patterns in terms of a feature vector. On the other hand, a set of dissimilarities may be more easily available as it is argued that the notion of proximity between patterns is more fundamental than that of features [23].

The dissimilarity representation of data is defined w.r.t. a "representation set"

$$R = \{p_1, p_2, \ldots, p_n\}$$

made up of n objects that are used as a reference for all other objects of interest. An object x can be represented in terms of dissimilarities as a vector

$$[d(x, p_1), d(x, p_2), \ldots, d(x, p_n)]$$

where $d(\cdot, \cdot)$ is a distance measure between pair of objects. This distance may be computed using some intermediate feature representation of patterns.

14.3.1 Dissimilarity Evaluation and Prototype Selection

The process usually employed by humans in assigning a class label to an object involves the use of some measure of "similarity" between objects. An object is thus assigned a label according to the class label of the most "similar" patterns whose label is known. Such measures of similarity may depend or not from some quantitative measure made on the objects. These measures are usually called "features." It is easy to see that the definition is, suitable features should depend on the notion of similarity between the objects belonging to the domain of interest. However, as in many applications it is hard to easily extract a set of effective features, it is common to extract a large number of candidate features, and then select the more significant subset for the task at hand. For this selection process to be effective the number of labeled prototypes should be large w.r.t. the number of extracted features. Unfortunately in many application domains the use of a large number of features is accompanied by a relatively small number of labeled prototypes. As a consequence, there are a number of difficulties in solving the pattern recognition problem by a statistical formulation. The dissimilarity representation of data aims at coming back to the roots of pattern recognition and machine learning, by emphasizing the role of similarity between patterns w.r.t. the feature representation of patterns [23, 26]. This focus on dissimilarity representation however does not exclude the use of feature spaces where patterns can be represented. On the other hand, it focus on formulating the problem in terms of dissimilarity between patterns, regardless the way such dissimilarities are computed [23]. Thus, dissimilarities between patterns may be computed using some feature spaces, but the problem itself is not formulated in some feature spaces, but in the dissimilarity space.

In the field of image retrieval for large image databases, usually a large number of low-level features are extracted as the semantic content of images typically exhibit a high variability (they are usually referred to "broad domain" database). In addition, as different users typically have different goals, the extraction of suitable features is not an easy task. To this end research in the field of CBIR focused on defining suitable techniques for manipulating the feature space (feature selection, feature weighting, feature space transformations, etc.). On the other hand, the use of a dissimilarity space may represent an alternative solution, as the low-level feature spaces can be used to compute similarities, which are then used to build a new space. Some researchers recently proposed to use a "manipulation space," i.e., a space where the dissimilarity between images are visualized in a 2D space [20]. In this way, the user may provide the feedback to the system by marking those images that are relevant to the query. However, it is worth noting that while in the feature space relevant images may be represented as "distant" points, in the dissimilarity space these points should be represented as"close" to each other. This effect can be explained by observing that two similar images should exhibit similar distances from some of the prototypes of the representative set. Thus the images are close each other in the dissimilarity space.

From the above discussion it is clear that a key role is played by the selection of the prototypes used to build the representative set [20, 23, 24]. It has been shown that in a number of pattern recognition applications the selection of prototypes can be performed randomly. On the other hand, a number of selection techniques can also be used for prototype selection. In the field of image retrieval, prototypes can be selected by browsing the image collection [20]. First images are clustered according to some criteria, and then one representative for each cluster are showed to the user. The user then selects the images that are more relevant to the query, and they are used as prototypes to build the dissimilarity space. Prototypes may also be chosen by employing the well-known editing techniques proposed both to speed-up nearest neighbor and case-based techniques, and to increase their performance by eliminating noisy or redundant cases [7,18]. However, as the use of nearest neighbor and case-based techniques for relevance feedback in image retrieval is still at an early stage, such techniques are worth to be considered for further improvements.

In conclusion, it can be said that typically dissimilarities are computed in some feature spaces, while prototypes may be chosen in a number of ways. The key concept in the dissimilarity representation is that two similar objects are close to each other in the dissimilarity space as they typically exhibit similar distance from at least some of the prototypes used to build the dissimilarity space.

The use of "manipulation spaces," where objects are displayed in a 2D space, may help in devising user-defined dissimilarities. The user may use this kind of visualization for "moving" relevant images close to each other. Then this visual movement should be transformed in quantitative dissimilarities to be used for further processing. These kinds of tools are currently at an early stage. For example, in [19] one of such tools have been proposed. In this case, the movement of images is used to compute weights for a weighted distance metric in some low-level feature space.

Thus, in the image retrieval field, the use of some low-level feature space is recognized as a useful tool for representing image content. However these features cannot be used directly to measure the similarity between images, but some processing is needed in order to bridge the gap between the low-level representation and the user perception of similarity. The dissimilarity representation is one approach that allows bridging the gap. This representation may also allow for using some user-defined similarity measure that does not depend on low-level feature space. However, the definition of such a measure is not an easy task, and it is argued that such a definition cannot be used as an alternative to low-level image representation, but as an additional measure to better estimate similarity between images.

14.3.2 The Proposed Dissimilarity Representation of Images for Relevance Feedback

In the proposed relevance feedback mechanism for image retrieval, we define the set R as the set made up of all the relevant images that the user has marked during the relevance feedback iterations. Thus the dimensionality of data strictly depends on the number of relevant images retrieved so far. In addition, the image representation depends on the user concept of similarity.

Let us consider an image database whose images I are represented in a d-dimensional low-level feature space, e.g., color, texture, etc. Let us assume that a dissimilarity metric $d(I_j, I_k)$ has been defined in such a feature space. In the following, we will neither make any assumption about the feature space, nor about the similarity metric employed. If R is the set of relevant images the user has marked during the first k relevance feedback iterations, and n is the size of R, the proposed dissimilarity representation I_{diss} of an image I from the database is the following:

$$I_{diss} = [d(I, I_{r,1}), d(I, I_{r,2}), \ldots, d(I, I_{r,n})] \tag{14.1}$$

where I_r represents an image belonging to R. It is worth noting that (14.1) is also used to represent the relevant images $I_1,\ I_2, \ldots,\ I_{Nr}$. This representation allows using the Euclidean distance measure to compute the dissimilarity between pairs of images [23].

The proposed dissimilarity representation strictly depends on the set of images that the user has marked to be relevant. Thus it can be argued that using this representation, an image I_{diss} will be as much as relevant as it is *near* to the relevant images and, at the same time, *far* from the nonrelevant ones.

In the following all the proposed formulas refer to the dissimilarity representation of images. Thus, for the sake of clarity, we will omit the subscript "*diss*."

14.4 Instance-Based Relevance Estimation

The proposed mechanism has been inspired by classification techniques based on the "nearest case," which are used in pattern recognition and machine learning for classification and outlier detection [1,2,4,9,36]. The present section illustrates the rationale behind the use of the nearest case paradigm, and provides the details of the techniques that we propose to measure the relevance of images.

Nearest neighbor techniques, as used in statistical pattern recognition, case-based reasoning, or instance-based learning, are effective in all applications where it is difficult to produce a high-level generalization of a "class" of objects. Relevance learning in content base image retrieval may well fit into this definition, as it is difficult to provide a general model that can be adapted to represent different concepts of similarity. In addition, the number of available cases may be too small to estimate the optimal set of parameters for such a general model. On the other hand, it can be more effective to use each "relevant" image as well as each "nonrelevant" image, as "cases" or "instances" against which the images of the database should be compared [14]. Consequently, we assume that an image is as much as relevant as much as its dissimilarity from the nearest relevant image is small. Analogously, an image is as much as nonrelevant as much as its dissimilarity from the nearest nonrelevant image is small. As this assumption may not hold in a feature space representation, images are represented in terms of dissimilarities, as illustrated in Sect.14.3.

14.4.1 Relevance Score Computation

The degree of relevance can be computed as follows. Let us recall that the nearest neighbor (NN) classifier is derived from the local estimation of densities in the neighborhood of the test pattern [9]. Such a local density can be written as [36]

$$p_{NN}(I) = \frac{1/N}{V\left(\|I - NN(I)\|\right)} \tag{14.2}$$

where N is the number of training patterns, I the test image, and NN denotes the nearest neighbor of I. Thus we can compute the local density of relevant images in I as

$$p^r_{NN}(I) = \frac{1/N}{V\left(\|I - NN^r(I)\|\right)} \tag{14.3}$$

where NN^r is the nearest relevant image of I. Analogously the local density of nonrelevant images can be computed as

$$p^{nr}_{NN}(I) = \frac{1/N}{V\left(\|I - NN^{nr}(I)\|\right)} \tag{14.4}$$

where NN^{nr} is the nearest nonrelevant image of I.

These densities can be used to estimate the degree of relevance of an image as

$$\begin{aligned} relevance(I) = P\left(relevant|I\right) &= \frac{p^r_{NN}}{p^r_{NN} + p^{nr}_{NN}} \\ &= \frac{\|I - NN^{nr}(I)\|}{\|I - NN^r(I)\| + \|I - NN^{nr}(I)\|} \end{aligned} \tag{14.5}$$

The relevance score computed according to (14.5) is then used to rank the images and the first k are presented to the user. It is worth noting that this relevance score can be thought of as an estimation of the posterior in I, as it is computed from an estimation of densities. However, as the estimation of densities cannot be deemed reliable as they are based on a very small training set, we will refer to this measure of relevance as a "relevance score."

14.4.2 Stabilization of the Relevance Score

The proposed score suffers from two problems. First of all, let us consider the training set size. Typically the number of images presented to the user by the retrieval system is in the order of few dozens (e.g., 20 images). When the user marks the first set of images retrieved by the system in response to the query, typically very few of them are relevant, the remaining being nonrelevant. It is easy to see that in these cases the density of relevant images computed according to (14.2) is small almost elsewhere, so that the proposed score output large values for those images *far* from the nonrelevant images (i.e., for those images where the density of nonrelevant images is small too).

To solve this problem, we propose to use the distance of I from a modified query vector computed according to [13]. This modified query vector is aimed at moving the search toward regions of the original feature space where it is more likely to find relevant images. We call this new query vector as “Bayesian query shifting” (BQS) as its formulations is derived from the Bayes decision theory:

$$Q_{BQS} = \mathbf{m}_R + \frac{\sigma}{\|\mathbf{m}_R - \mathbf{m}_N\|}\left(1 - \frac{k_R - k_N}{\max\left(k_R, k_N\right)}\right)(\mathbf{m}_R - \mathbf{m}_N) \qquad (14.6)$$

where $\mathbf{m}_R$ and $\mathbf{m}_N$ are the mean vectors of relevant and nonrelevant images, respectively, $\boldsymbol{\sigma}$ is the standard deviation of the images belonging to the neighborhood of the original query, and k_R and k_N are the number of relevant and nonrelevant images, respectively. The new query Q_{BQS} lies on the line connecting the two means, in the $\mathbf{m_R} - \mathbf{m_N}$ direction, the magnitude of the shift depending on the proportion of relevant and nonrelevant images retrieved. It is easy to see that the larger the number of nonrelevant images retrieved, the larger the magnitude of the shift. For more details about this technique, the reader is referred to [13]. It has been shown that this new query vectors allows attaining good performances in terms of retrieval precision, especially when the number of relevant images is small.

Let us denote with d_{BQS} the distance of image I from Q_{BQS}

$$d_{BQS}(I) = \|I - Q_{BQS}\| \qquad (14.7)$$

In order to combine this distance with the relevance score, we need to transform the distance into a score in [0,1]. We used a Gaussian model and denoted the resulting score as *locality(I)*:

$$locality(I) = \frac{1 - e^{1 - d_{BQS}(I) \Big/ \max\limits_{I} d_{BQS}(I)}}{1 - e} \qquad (14.8)$$

In order to compute the stabilized score, let r and n be the number of relevant and nonrelevant images retrieved after the latter iteration, respectively. The stabilized score can be computed as follows:

$$relevance(I)_{stab} = \left(\frac{n/k}{1+n/k}\right) \cdot locality(I) + \left(\frac{1}{n/k+1}\right) \cdot relevance(I) \tag{14.9}$$

The weights of the combination are both equal to 1/2 when no relevant image is retrieved in the latter iteration, while the weights of *locality(I)* decreases as the number of relevant images increases. The weights of *locality(I)* goes to zero when all the retrieved images are relevant.

The second problem in nearest neighbor density estimation is related to the reliability of the estimation with small sample size. It is worth recalling that the previous issue was related to the cases when the number of relevant images is small, while in this paragraph we are addressing the issue of small training set size, i.e., the fact that the number of retrieved images is small. In these cases the use of the first nearest neighbor can hardly be considered a reliable estimation of the local density. In particular this problem is more severe near the boundary between relevant and nonrelevant images. To solve this problem, we propose to use the kth distance instead of the distance from the first nearest neighbor in (14.3)–(14.5) [4, 36].

14.5 Experimental Results

In order to test the proposed method and compare it with other methods described in the literature we used a subset of the Corel data set (Fig. 14.1). This data set is currently used for assessing and comparing relevance feedback techniques.

The data set extracted from the Corel collection is available at the KDD-UCI repository (http://kdd.ics.uci.edu/databases/CorelFeatures/CorelFeatures.data.html). We used a subset made up of 19,511 images, manually

Fig. 14.1. Some examples of the images contained in the Corel data set

subdivided into 42 semantic classes. For each image, the four sets of features available at the Web site have been considered, i.e., Color Histogram (32 features), Color Histogram Layout (32 features), Color Moments (nine features), and Cooccurrence Texture, (16 features). More details on the feature extraction process can be found in [21]. In particular, the similarity between pairs of images is computed according to the Manhattan distance for the first two sets of features, while the Euclidean distance is used for the latter two sets of features. A linear normalization procedure has been performed, so that each feature takes values in the range between 0 and 1.

500 images have been randomly extracted and used as query. The top 20 nearest neighbors of each query 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 20 as nonrelevant. This experimental set up affords an objective comparison among different methods and is currently used by many researchers. Performances are computed in terms of the retrieval precision, i.e., the average percentage of relevant images among the top 20 images retrieved by the systems.

The proposed relevance feedback technique has been implemented by considering the second-NN distance. In particular, we report results attained by the *relevance* score computed according to (14.5), and the ones attained by the *stabilized relevance* score computed according to (14.9). Thus it is possible to compare the behavior of the "pure" instance-based technique (14.5) with the instance-based technique combined with the *locality* term (14.9).

For the sake of comparison, retrieval performances obtained with three methods recently described in the literature, are also reported, namely the BQS and SVM. The BQS technique has been illustrated in Sect. 14.4 as it is also used in the present chapter in combination with the instance-based rule. The SVM have been trained using the sets of relevant and nonrelevant images as a training set of a two-class classification problem, and images have been ranked according to the SVM output. As SVM training requires choosing the kernel and the learning parameters, we used two commonly used kernels, namely the linear and the Gaussian kernel. In particular, for the SVM with Gaussian kernel, reported results are the best ones attaining with different values of the learning parameters, while in the case of linear kernel no parameter optimization is required.

Figures 14.2–14.5 show the results attained with the four sets of features of the Corel data set. The retrieval performances attained after the query image is presented to the system are quite low, thus showing that the chosen feature sets are not suited for the task at hand. In addition, different sets of features provided different results, thus confirming that the choice of the set of features is a key aspect of CBIR systems.

It is easy to see that each of the considered relevance feedback techniques allow improving the retrieval precision. The only exception is the linear SVM, which is not suited for the Color Histogram and Color Histogram Layout sets of features. In fact the retrieval performances not only does not improve with

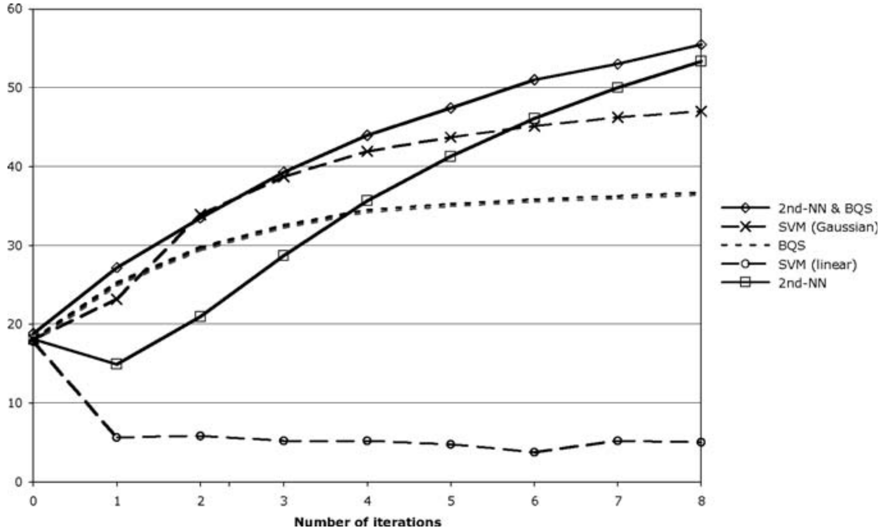

Fig. 14.2. Average percentage precision for the Color Histogram feature set. Eight relevance feedback iterations were performed

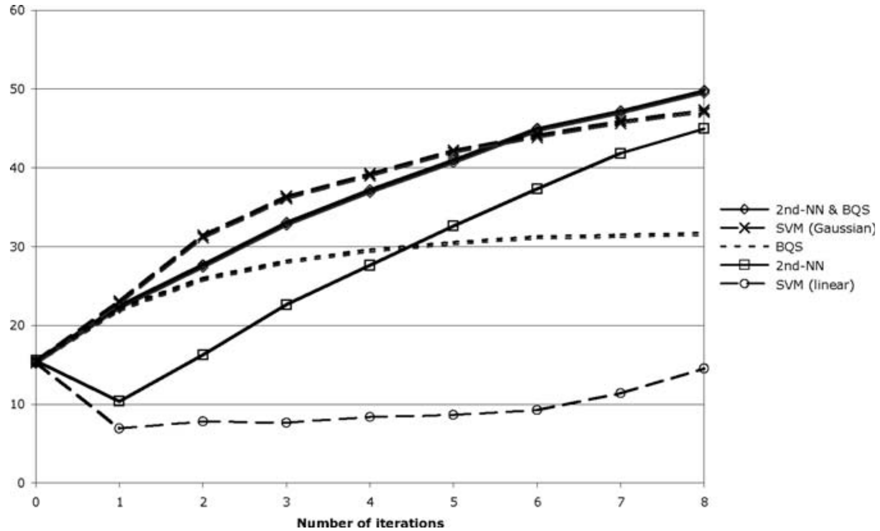

Fig. 14.3. Average percentage precision for the Color Histogram Layout feature set. Eight relevance feedback iterations were performed

respect to the first retrieval, but also they get worse. On the other hand, linear SVM provided performance improvements on the other two sets of features.

The proposed instance-based technique in the dissimilarity space combined with BQS, provided performances higher than those provided by the SVM with Gaussian kernel. This superiority is more evident in Figs. 14.3–14.5. It is

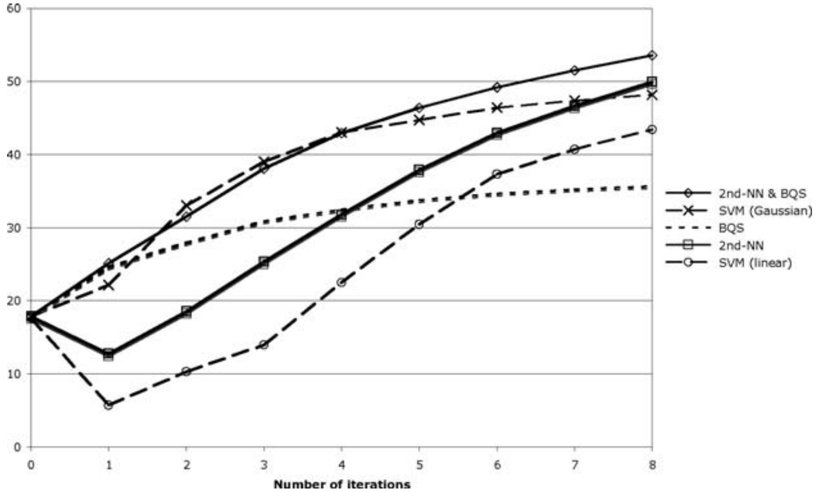

Fig. 14.4. Average percentage precision for the Color Moments feature set. Eight relevance feedback iterations were performed

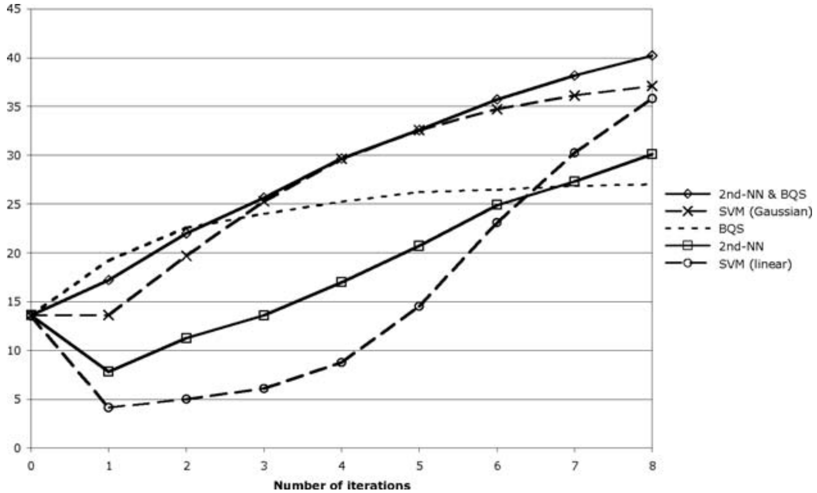

Fig. 14.5. Average percentage precision for the Cooccurrence texture feature set. Eight relevance feedback iterations were performed

worth recalling that the reported performances of the SVM with Gaussian kernel are the best one attained with different parameter values. On the other hand, the proposed mechanism does not require setting any parameter. Thus the proposed technique allows attaining good performances without requiring parameter tunings.

If we analyze the results, we see that the precision decreases in the first iteration when the “pure” second-NN technique, and the linear SVM are used.

This effect is motivated by the fact the in the first iteration very few relevant images are usually retrieved. As a result SVM and second-NN assign high scores to images that are *different* form the nonrelevant images. As the problem is not a two-class problem but a $(1 + x)$ problem, thus it follows that an image that is different from some nonrelevant images is not necessarily a relevant image. If the iterations following the first are considered, the precision attained by the second-NN technique increases in all of the four considered sets of features. In the case of the linear SVM, the precision increases only in two out of the four sets of features, namely Color Moments and Cooccurrence texture.

On the other hand, the performances of BQS allows attaining high performance increases in the first iteration, while the increase is modest in the following iterations. This behavior may be due to the locality of BQS that does not allow exploring new regions of the feature space.

Gaussian SVM and the stabilized second-NN (second-NN & BQS) provide the highest performances. In all the considered feature sets the precision of the stabilized second-NN is always higher than the one provided by the Gaussian SVM. The only exceptions are the iterations from 2 to 5 in the Color Histogram Layout feature set, and the iterations 2 and 3 in the Color Moments feature set where the performances of the Gaussian SVM are slightly higher than those provided by the stabilized second-NN.

However, it is worth recalling that reported results for the Gaussian SVM are related to the best one attained in a number of trials. In a number of experiments with different values of the learning parameters, the performances of the Gaussian SVM after the first relevance feedback iteration were usually smaller than those attained in the first retrieval (i.e., the k nearest neighbors of the query). As the estimation of the optimal or suboptimal parameters for the Gaussian SVM in relevance feedback is beyond the state of the art, it can be concluded that the proposed mechanism is a robust technique in different image representation contexts.

14.6 Conclusions

The dissimilarity representation of data is receiving increasing attention in a number of applications. Recent research in relevance feedback for CBIR proposed some techniques based on this representation of data. This representation allows for avoiding the direct use of low-level feature spaces that may not be suited to the user's needs.

The dissimilarity representation proposed in this chapter use the set of relevant images retrieved so far as the representation set. Then, images are ranked according to a relevance score computed by a combination of two terms, namely a *relevance* term based on the distances from the second relevant neighbor and the second nonrelevant neighbor, and a *locality* term computed in terms of a shifted query in the original feature space. These two terms plays the role of an exploration term and exploitation term, respectively.

Reported results on four feature sets of the Corel image database showed the superiority of the proposed method with respect to state-of-the-art relevance feedback techniques. In particular it has been pointed out that the proposed technique does not require any parameter setting, while other relevance feedback techniques require a long phase for tuning the parameters. Thus it can be concluded that the proposed relevance feedback mechanism is a robust tool that allows attaining good performances in a number of different image representations.

As far as the computational complexity of the proposed technique is concerned, a large number of distances are to be computed. Nevertheless, the response time between two consecutive feedbacks is far below the classic limit of 1.0s for the user's flow of thought to stay uninterrupted. Thus, despite the computational complexity of the algorithm, the response time on a typical PC configuration can be considered acceptable for a large database. However, the response time of the implemented algorithm could be further improved by using some editing techniques for decreasing the number of distances to be evaluated.

Acknowledgments

This work was partially supported by the Italian Ministry of University and Research within the framework of the project "Similarity-based Methods for Computer Vision and pattern recognition: Theory, Algorithms, Applications."

The authors wish to thank Dr. Michael Ortega for providing the images of the Corel Data set, and Dr. Petra Perner for her useful comments and suggestions.

References

1. Aha DW, Kibler D, Albert MK (1991) Instance Based learning Algorithms. Machine Learning 6:37–66
2. Althoff K-D (2001) Case-Based Reasoning. In Chang S.K. (ed.) Handbook on Software Engineering and Knowledge Engineering, World Scientific, 549–588
3. Bhanu B, Dong D (2001) Concepts Learning with Fuzzy Clustering and Relevance Feedback. In: Perner, P. (Ed.): Machine Learning and Data Mining in Pattern Recognition. LNAI 2123, Springer-Verlag, Berlin 102–116
4. Breunig M, Kriegel H-P, Ng R, Sander J. (2000) LOF: indentifying density-based local outliers. In Proc. of the ACM SIGMOD 2000 Int. Conf. on management of data
5. Bruno E, Loccoz N, Maillet S (2005) Learning user queries in multimodal dissimilarity spaces. Proc. of the 3rd Int'l Workshop on Adaptive Multimedia Retrieval

6. Cox I.J, Miller ML, Minka TP, Papathomas TV, Yianilos PN (200) The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments. IEEE Trans. on Image Processing 9:20–37
7. Dasarathy DV (Ed.) (1991) Nearest Neighbor Norms: NN Pattern Classification Techniques. IEEE Press
8. Del Bimbo A (1999) Visual Information Retrieval. Morgan Kaufmann Pub. Inc., San Francisco, CA
9. Duda RO, Hart PE, Stork DG (2001) Pattern Classification. John Wiley and Sons, Inc., New York
10. Duin RPW, de Ridder D, Tax DMJ (1997) Experiments with object based discriminant functions: a featureless approach to pattern recognition. Pattern Recognition Letters 18:1159–1166
11. Frederix G, Caenen G, Pauwels EJ (2000) PARISS: Panoramic, Adaptive and Reconfigurable Interface for Similairty Search. Proc. of ICIP 2000 Intern. Conf. on Image Processing. WA 07.04, vol. III, 222–225
12. Giacinto G, Roli F (2003) Dissimilarity Representation of Images for Relevance Feedback in Content-Based Image Retrieval. In: Perner P. (Ed.) Machine Learning and Data Mining in Pattern Recognition. LNAI 2734, Springer-Verlag, Berlin 202–214
13. Giacinto G, Roli F (2004) Bayesian Relevance Feedback for Content-Based Image Retrieval. Pattern Recognition 37:1499–1508
14. Giacinto G, Roli F (2005) Instance-Based Relevance Feedback for Image Retrieval. In Saul L.K., Weiss Y., and Bottou L.: Advances in Neural Information Processing Systems 17, MIT Press 489–496
15. Ishikawa Y, Subramanys R, Faloutsos C (1998) MindReader: Querying databases through multiple examples. In Proceedings. of the 24th VLDB Conference 433–438
16. Lew MS, Sebe N, Djeraba C, Jain R (2006) Content-Based Multimedia Information Retrieval: State of the Art and Challenges. ACM Trans. On Multimedia Computing, Communications and Applications 2:1–19
17. McG Squire D, Müller W, Müller H, Pun T (2000) Content-based query of image databases: inspirations from text retrieval. Pattern Recognition Letters 21:1193–1198
18. McKenna E, Smyth B (2000) Competence-Guided Case-Base Editing Techniques in Blanzeri and Portinale (Eds.) Proceedings of the 5th European Workshop on Advances in Case-Based Reasoning, LNCS 1898, Springer-Verlag: 186–197
19. Moghaddam B, Tian Q, Lesh N, Shen C, Huang TS (2004) Visualization and User-Modeling for Browsing Personal Photo Libraries. International Journal of Computer Vision 56:109–130
20. Nguyen GP, Worring M, Smeulders AWM (2006) Similarity learning via dissimilarity space in CBIR. Proc. of the 8th ACM Int'l workshop on Multimedia Information retrieval 107–116
21. Ortega M, Rui Y, Chakrabarti K, Porkaew K, Mehrotra S, Huang TS (1998) Supporting ranked boolean similarity queries in MARS. IEEE Trans. on KDE 10:905–925
22. Pekalska E, Duin RPW (2002) Dissimilarity representations allow for building good classifiers. Pattern Recognition Letters 23:943–956
23. Pekalska E, Duin RPW (2005) The dissimilarity representation for pattern recognition: foundations and applications. World Scientific Publishing

24. Pekalska E, Duin RPW, Paclick (2006) Prototype selection for dissimilarity-based classifiers. Pattern Recognition 39:189:208
25. Peng J, Bhanu B, Qing S (1999) Probabilistic feature relevance learning for content-based image retrieval. Computer Vision and Image Understanding 75:150–164
26. Perner P (2002) Are case-based reasoning and dissimilarity-based classification two sides of the same coin? Engineering Applications of Artificial Intelligence 15:193–203
27. Rui Y, Huang TS, Mehrotra S (1997) Content-based image retrieval with relevance feedback in MARS. In Proceedings of the IEEE International Conference on Image Processing, IEEE Press 815–818
28. Rui Y, Huang TS (2001) Relevance Feedback Techniques in Image retrieval. In Lew M.S. (ed.): Principles of Visual Information Retrieval. Springer-Verlag, London, 219–258
29. Salton G, McGill MJ (1998) Introduction to modern information retrieval. McGraw-Hill, New York
30. Santini S, Jain R (1999) Similarity Measures. IEEE Trans. on Pattern Analysis and Machine Intelligence 21:871–883
31. Santini S, Jain R (2000) Integrated browsing and querying for image databases. IEEE Multimedia 7:26–39
32. Sclaroff S, La Cascia M, Sethi S, Taycher L (2001) Mix and Match Features in the ImageRover search engine. In Lew M.S. (ed.): Principles of Visual Information Retrieval. Springer-Verlag, London 219–258
33. Smeulders AWM, Worring M, Santini S, Gupta A, Jain R (2000) Content-based image retrieval at the end of the early years. IEEE Trans. on Pattern Analysis and Machine Intelligence 22:1349–1380
34. Tao D, Tang X, Li X, Rui Y (2006) Direct Kernel Biased Discriminant Analysis: A New Content-based Image Retrieval Relevance Feedback Algorithm. IEEE Trans. on Multimedia 8:716–727
35. Tao D, Tang X, Li X, Wu X (2006) Asymmetric Bagging and Random Subspace for Support Vector Machines-Based Relevance Feedback in Image Retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence 28:1088–1099
36. Tax D (2001) One-class classification. PhD thesis, Delft University of Technology, The Netherlands
37. Tieu K, Viola P (2001) Boosting Image Retrieval. Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, vol. 1, 228–235
38. Tong S, Chang E (2001) Support Vector Machine Active Learning for Image Retrieval. Proc. ACM Int'l Conf. Multimedia 107–118
39. Zhang L, Lin F, Zhang B (2001) Support Vector Machine Learning for Image Retrieval. Proc. IEEE Int'l Conf. Image Processing 721–724
40. Zhou X, Huang TS (2001) Small Sample Learning During Multimedia Retrieval Using Biasmap. Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, vol. 1, 11–17
41. Zhou X, Huang TS (2003) Relevance Feedback for Image Retrieval: A Comprehensive Review, ACM Multimedia Systems 8:536–544