

# Interactive Remote Sensing Image Retrieval Using Active Relevance Feedback

Marin Ferecatu and Nozha Boujemaa

**Abstract**—As the resolution of remote sensing imagery increases, the full complexity of the scenes becomes increasingly difficult to approach. User-defined classes in large image databases are often composed of several groups of images and span very different scales in the space of low-level visual descriptors. The interactive retrieval of such image classes is then very difficult. To address this challenge, we evaluate here, in the context of satellite image retrieval, two general improvements of the SVM-based relevance feedback method. First, to optimize the transfer of information between the user and the system, we focus on the criterion employed by the system for selecting the images presented to the user at every feedback round. We put forward an active learning selection criterion that minimizes redundancy between the candidate images shown to the user. Second, for image classes spanning very different scales in the low level description space, we find that a high sensitivity of the SVM to the scale of the data brings about a low retrieval performance. We argue that insensitivity to scale is desirable in this context and we show how to obtain it by the use of specific kernel functions. Experimental evaluation of both ranking and classification performance on a groundtruth database of satellite images confirms the effectiveness of our approach.

**Index Terms**—Image retrieval, sample selection, active learning, reduction of redundancy, kernel function.

## I. INTRODUCTION

EVERY day, large quantities of information are becoming available in remote sensing data repositories, bringing with them an increasing need for intelligent data access. Recently, the problem of retrieval from large unstructured remote sensing image databases has begun to be studied, boosted by the need to access relevant information in an understandable and directly usable form and to provide friendly interfaces for information query and browsing [1].

Remote sensing databases are operated by specialists from widely varying fields and the data access by time of acquisition, geographical position or type of sensor is often less important than the content of the scene, e.g. structures and objects. The needs of the user are thus precise and complex, yet very different from one application to another. Interesting applications involve complicated spatial and structural relationships among image objects [2]: high level elements such as buildings, bridges, human activity, and so on, are better characterized by their inter-relations than by their individual characteristics.

Thus, an important topic in satellite image content extraction and classification is building retrieval systems that automatically learn high-level semantic interpretations from images,

possibly under the direct supervision of the user. Traditional methods use local pixel characteristics to create a link between the low level features of the images and the high level descriptions needed by the users. For example, [3] propose a system that uses Bayesian classifiers to represent high-level land cover labels for pixels using their low-level spectral and textural attributes. The classifiers are used to retrieve images from remote sensing archives by approximating the probabilities of images belonging to different classes using pixel level probabilities. In [4], the authors use a hierarchical set of image segmentations and present a relevance feedback method that allow for automatic selection of features for region and tile similarity searches. Selim et al. [5] introduce a Bayesian framework for a visual grammar that aims to reduce the semantic gap by modeling image pixels using automatic fusion of their spectral, textural and other attributes. Parulekar et al. [6] use semantic categorization done by a learning approach involving the two-dimensional multi-resolution hidden Markov model. User's feedback may be also be used to automatically update the weights of the low level features, in an effort to isolate regions of the description space that can be associated with high level concepts [7].

In this paper we introduce two improvements of the SVM-based relevance feedback scheme and evaluate them on a database of satellite images. We use global image descriptors based on the statistical description of the most important visual characteristics (color, texture and shape) and we present several search examples which show that relevance feedback can provide very effective results on remote sensing databases.

The concept of “semantic gap” has been extensively used to express the discrepancy between the low-level features that can be readily extracted from the images and the high level descriptions that are meaningful to the users of a search engine. An effective solution for reducing the semantic gap is to cut the search session into several consecutive retrieval rounds and let the user provide feedback regarding the results of every retrieval round, e.g. by qualifying images returned as either “relevant” or “irrelevant” (Relevance Feedback: RF). From this feedback, the engine progressively learns the visual features of the images the user is looking for in the current search session (the “target” of the user).

The RF method embodied in a search engine should operate in real time and should maximize the ratio between the quality (or relevance) of the results and the amount of interaction between the user and the system. An RF method [8] is usually defined by two components: a learner and a selector. At every feedback round, the learner uses the images marked as “relevant” or “irrelevant” by the user to re-estimate the target

of the user. Given the current estimation of the target, the selector chooses the images for which the user is asked to provide feedback during the next round. Much recent work is based on support vector machines [9], [10] because they avoid too restrictive assumptions regarding the data (e.g. that classes should have elliptic shape), are very flexible (can be tuned by kernel engineering) and allow fast learning and evaluation for medium-sized databases.

The task of the learner is very difficult in the context of RF [11], [8], since training examples are scarce (their number is usually lower than the number of dimensions of the description space), the training set is heavily imbalanced (there are often many more “irrelevant” examples than “relevant” ones) and both training and evaluation must be performed in real time.

We would like to emphasize the fact that RF is usually applied to *two important scenarios* of different nature. The *first* and most common one is finding images in a specific target set: the focus is on ranking most of the “relevant” images before the “irrelevant” ones. The *second* use of RF is for defining a class of images for later use: in this case the focus is on identifying a good frontier between the class of interest and the other images.

In much of the work on RF, the images for which the user is asked to provide feedback at the next round were simply those that were currently considered by the learner as the most relevant; also, in some cases these images are randomly selected. An important step ahead was the introduction in [12] and [13] of an *active learning* framework for RF using SVMs. We put forward here an improvement of this framework, consisting of a selection criterion that reduces the redundancy between the images for which the user is asked to provide feedback; our criterion encourages the selection of images that are far from each other in the space of low-level visual descriptors and thus allows for a better exploration of the current frontier.

The image classes (that is, the potential targets of the users of an RF system) can have various shapes and span very different scales in the space of low-level visual descriptors. Learners that are strongly dependent on a scale parameter will then be unable to find for this parameter a value that is adequate for all the classes in a database; with very few training examples, the scale parameter is difficult to tune online to suit the current class. We argue that a low sensitivity of the learner to scale of the data is then an important desirable feature in an RF context and we propose to use specific kernel functions that let SVMs achieve this.

The paper is organized as follows: in Section II we describe our active selection criterion with reduction of redundancy, in Section III we discuss the issue of the insensitivity of the SVM to the scale of the data and in Section IV we present experimental evidence that our relevance feedback mechanism can be applied with excellent results for querying remote sensing image databases.

## II. REDUCTION OF THE REDUNDANCY

In order to maximize the ratio between the quality (or relevance) of the results and the amount of interaction between

the user and the system, the selection of images for which the user is asked to provide feedback at the next round must be carefully studied.

Interesting ideas were introduced in [14], [15], where the problem under focus is the iterative search for one specific image in a database (*target* search); at every round, the user is required to choose, between the two images presented by the engine, the one that is closest to the target image. The selection criterion put forward in this case attempts to identify at every round the most informative binary selections, i.e. those that are expected to maximize the transfer of information between the user and the engine (or remove a maximal amount of uncertainty regarding the target). We consider that this criterion translates into two complementary conditions for the images in the selection: each image must be ambiguous given the current estimation of the target and the redundancy between the different images has to be low.

Unfortunately, the entropic criterion employed in [14], [15], does not scale well to the search of images in a larger set (*category* search) and to the selection of more than 2 images. Computational optimizations must be found, relying on the use of specific learners and, possibly, specific search contexts.

Based on the definition of active learning [16], the selection of examples for training SVMs to perform general classification tasks is studied in [17]. When the classification error increases with the distance between the misclassified examples and the frontier (a “soft margin” is used for the SVM), the authors interestingly distinguish two cases: early and late stages of learning.

In the early stages, the classification of new examples is likely to be wrong, so the fastest reduction in generalization error can be achieved by selecting the example that is farthest from the current estimation of the frontier. During late stages of learning, the classification of new examples is likely to be right but the margin may be suboptimal, so the fastest reduction in error can be achieved by selecting the example that is closest to the current estimation of the frontier. Note that, according to the classical formulation of active learning, the authors only consider the selection of a single example for labeling (for addition to the training set) at every round.

For SVM learners, several selection criteria are presented in [12] and applied to content-based text retrieval with relevance feedback. The simplest of these criteria consists of selecting the document that is closest to the hyperplane currently defined by the SVM. We call this criterion the selection of the “most ambiguous” (MA) candidate. This selection criterion is motivated by the need to choose query instances that split the current version space into two equal parts as much as possible. In this case, the version space is the set of parameters of the hyperplanes in feature space that are compatible with the already labeled examples. This result assumes that the version space is not empty and that, in the feature space associated to the kernel, all the images of vectors in the input space have constant norm. These assumptions hold with appropriate choices for the kernel and for the upper bound  $C$  on the Lagrange multipliers of the SVM’s decision function. The MA criterion relies on the assumption that the version space is fairly symmetric, which is not always the case, but it has

the big advantage of being fast. This is an important factor for image retrieval, since relevance feedback should return answers in real time for fairly large databases.

In [13] the MA selection criterion is applied to CBIR with relevance feedback and shown to produce a faster identification of the target images than the selection of random images for labeling. In order to minimize the number of learning rounds, the user is asked to label several examples at every round and all these examples are selected according to the MA criterion.

While the MA criterion provides a computationally effective solution to the selection of the most ambiguous images, when used for the selection of more than one candidate image it does not remove the redundancies between the candidates.

We propose here to introduce the following additional *condition of low redundancy*: if  $x_i$  and  $x_j$  are the input space representations of two candidate images, then we require a low value for  $K(x_i, x_j)$ , which is the value taken by the kernel for this pair of images. If the kernel  $K$  is inducing a Hilbert structure on the feature space, if  $\phi(x_i)$ ,  $\phi(x_j)$  are the images of  $x_i$ ,  $x_j$  in this feature space and if all the images of vectors in the input space have constant norm, then this additional condition corresponds to a requirement of (quasi-)orthogonality between  $\phi(x_i)$  and  $\phi(x_j)$  (since  $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$ ). We shall call this criterion the selection of the “most ambiguous and orthogonal” (MAO) candidates.

We note that the MAO criterion can be extended to reduce redundancies between the examples selected during subsequent RF rounds. This additional constraint may be important in situations where the number of labeled examples is much lower than the dimension of the input space and the classes are restricted in most directions.

The MAO criterion has a simple intuitive explanation for kernels  $K(x_i, x_j)$  that decrease with an increase of the distance  $d(x_i, x_j)$ , which is the case for most common kernels: it encourages the selection of unlabeled examples that are far from each other in the input space, allowing to better explore the current frontier.

To implement this criterion, we first perform an MA selection of a larger set of unlabeled examples. Then, we build from it an MAO selection by iteratively choosing as a new example the vector  $x_j$  that minimizes the highest of the values taken by  $K(x_i, x_j)$  for all the  $x_i$  examples already included in the current MAO selection. This can be written as:

$$x_j = \operatorname{argmin}_{x \in S} \max_i K(x, x_i) \quad (1)$$

where  $S$  is the set of images not yet included in the current MAO selection and  $x_i$ ,  $i = 1 \dots n$  are the already chosen candidates.

This condition is justified by reference to the version space account suggested in [12]: diversity is maximized when the hyperplanes associated to the individual examples are orthogonal and are thus complementary to each other in halving the version space.

In a general classification context, [18] proposed the use of a trade-off between choosing non-redundant candidates and choosing candidates that are close enough to the hyperplane boundary. However, this criterion is computationally expensive and does not scale well to large image databases. Since

most of the images are situated far away from the SVM frontier, it is not necessary to optimize the low redundancy condition over a large set of images. Choosing candidates from a small pre-selection around the frontier, as we propose here, is much faster and does not have a significant impact on the results. Our criterion is as fast as MA [13] while eliminating the redundancy between the selected samples: the overhead imposed is  $O(\text{constant})$ , independent of the size of the database, unlike the algorithm proposed in [18].

We note that the MA criterion in [12] and [13] is the same as the one put forward in [17] for the late stages of learning. This clarifies the fact that the MA criterion relies on two important further assumptions: *first*, the prior on the version space is rather uniform; *second*, the solution found by the SVM is close to the center of gravity of the version space. The second assumption can be relieved by using Bayes Point Machines [19] instead of SVMs or the more sophisticated criteria put forward in [12], albeit at a higher computational cost.

### III. INSENSITIVITY TO SCALE

During the study of several groundtruth databases we found that, in the space of visual descriptors, the size of the various image classes often covers an important range of different scales (1 to 7 in our tests). We expect yet more significant changes in scale to occur from one database to another, from one user-defined class to another within a large real-world image database and even between parts of the frontier of some classes. A too strong sensitivity of the learner to the scale of the data could then strongly limit its applicability in an RF context.

For SVM classifiers, sensitivity to scale has two sources: the scale parameter of the kernel and the upper bound  $C$  on the Lagrange coefficients of the SVM. We focus here on the first source of sensitivity, the second one being usually less constraining (the bound  $C$  can be set in our retrieval context to some high value without significantly affecting performance).

The first kernel we consider is the Gaussian (RBF) one,  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ . This kernel is highly sensitive to the scale parameter  $\gamma$  (the inverse of the variance of the Gaussian).

The use of the Laplace (or exponential) kernel,  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|)$ , was advocated in [20] for histogram-based image descriptors. In [21], this kernel was found to work better than the Gaussian kernel for CBIR with RF.

The hyperbolic kernel,  $K(x_i, x_j) = 1/(\varepsilon + \gamma \|x_i - x_j\|)$ , can be computed fast and we have already used it for RF with good results. The scale parameter is  $\gamma$  again;  $\varepsilon$  translates into a multiplicative constant plus a change in  $\gamma$  and is only used to avoid numerical problems (we set it to 0.001).

All the three kernels we mentioned are positive definite kernels. The triangular kernel,  $K(x_i, x_j) = -\|x_i - x_j\|$ , was introduced in [22] as a *conditionally* positive definite kernel, but the convergence of SVMs remains guaranteed with this kernel [23]. In [24] the triangular kernel was shown to have a very interesting property: it makes the frontier found by SVMs invariant to the scale of the data.

Although particularly important for relevance feedback, the scale issue has rarely been mentioned explicitly in the litera-

ture. We suggest here that kernels inducing a low sensitivity of the learner to the scale of the data, such as the triangular kernel presented above, should be preferred with relevance feedback. Since the user-defined classes are not known *a priori*, the scale-invariance obtained by the use of the triangular kernel becomes a highly desirable feature and, according to the experimental evidence (see Sec. IV), makes this kernel an excellent choice.

Since the triangular kernel is not positive definite but only conditionally positive definite, the account provided in [12] and [18] for the MA selection criterion does not hold for this kernel. For the same reason, the triangular kernel does not induce a Hilbert structure on the feature space, so in this case one should not speak of the “orthogonality” of vectors in the feature space. However, since the value of  $K(x_i, x_j)$  decreases with an increase of the distance  $d(x_i, x_j)$ , our justification for the MAO criterion holds, as well as the justification of the MA criterion in [17].

#### IV. EXPERIMENTAL EVALUATION

In this section we present experimental evidence, using a groundtruth database of satellite images, that confirms the effectiveness of our propositions. We also present some sample retrieval screenshots both in a query by example context and with relevance feedback and we argue that relevance feedback can be a very effective solution for mining large remote-sensing image databases.

##### A. Groundtruth database and image descriptors

A specific RF method can be developed and evaluated on a particular application, with a well-defined scenario and a well-identified group of users. Knowing the specific assumptions concerning the application, the scenario and the users may help optimizing the RF method. It is nevertheless important to find improvements to RF methods that are relatively general and apply to many contexts. Evaluating such improvements by experimenting with human users is very difficult to set up, since it would require the cooperation of many different groups of people in various contexts. The common alternative is to use image databases for which a ground truth is available; this ground truth usually corresponds to the definition of a set of mutually exclusive image classes, covering the entire database.

**Groundtruth database.** We use a groundtruth database of satellite images build by hand at TSI ENST Paris for the QuerySat Project<sup>1</sup>, from a large satellite image database kindly provided by the Institut Gographique National, France<sup>2</sup>. The groundtruth database has 3600 images in 6 classes, each class having 600 images. Each class contains typical satellite images illustrating a category that is likely to be searched for by a real user: “city”, “cloud”, “desert”, “field”, “forest” and “sea”. As an illustration, Fig 1 and Fig 2 show some sample images from the classes “city” and “clouds” respectively.

**Image descriptors.** To describe the visual content of the images, we employ weighted histograms described in [25],

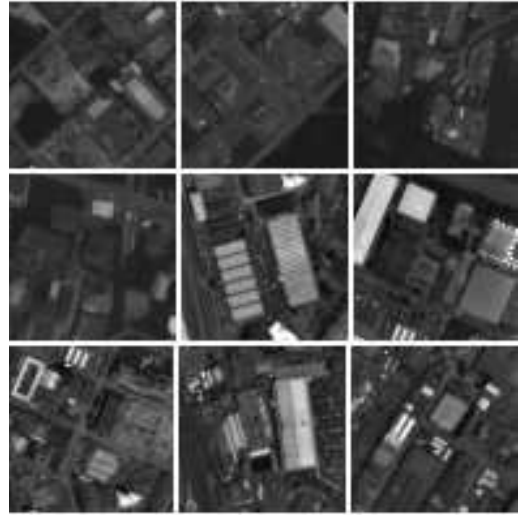


Fig. 1. Groundtruth database: samples images from the class “city”.

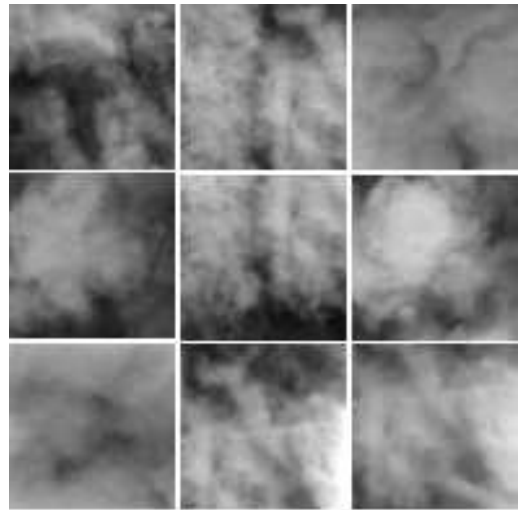


Fig. 2. Groundtruth database: samples images from the class “clouds”.

[26] using the Laplacian and the local probability of colour as pixel weighting functions. Weighting functions bring additional information into the histograms (e.g. local shape or texture), which is an important principle in building reliable image descriptors. The resulting integrated image descriptors generally perform better than a combination of classical, single-aspect features. Moreover, weighted histograms work equally well for color images and for gray level images, such as those in our test database of satellite images.

To describe the shape content of an image we use a histogram based on the Hough transform, which gives the global pixel behavior along straight lines in different directions. Texture feature vectors are based on the Fourier transform, yielding a distribution of the spectral power density along the frequency axes. This image descriptor performs well on texture images and, used in conjunction with other image descriptors, can significantly improve the overall behavior. For a detailed presentation of the image features see [26] [27]. The resulting joint feature vector has more than 600 dimensions.

<sup>1</sup><http://www.tsi.enst.fr/QuerySat/index.html>

<sup>2</sup><http://www.ign.fr>

**Dimensionality reduction.** The very high number of dimensions of the joint feature vector can make RF impractical even for medium-size databases. Also, the higher the dimension of the description space, the more difficult is the task of the learner. In order to reduce the dimension of the feature vectors, we use linear principal component analysis (PCA), applied separately to each of the image features previously described (see Fig 3).

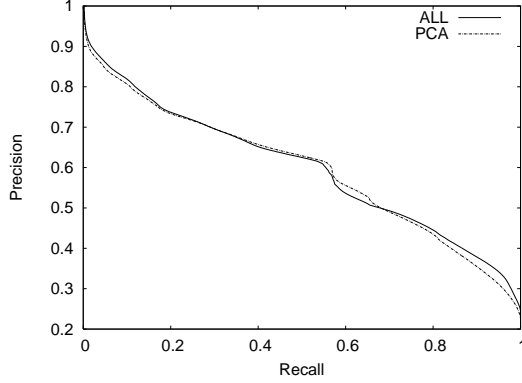


Fig. 3. Precision/recall diagram: after a reduction in dimension using linear PCA of about 5 times, we obtain a less than 3% overall loss of precision for any recall value.

We expected kernel PCA [10] to better focus on relevant nonlinear “dimensions”; this should indeed be the case when the manifold spanned by the images is very low-dimensional but significantly nonlinear. However, when comparing KPCA to linear PCA we noticed that the first did not perform so well on the image database we are using, suggesting that the previous assumption is wrong in these case.

A good criterion for choosing the number  $k$  of dimensions to keep after the PCA is, for example, that the energy of eliminated dimensions is less than 1% of the total energy in all dimensions.

$$\frac{\sum_{i=k+1}^N \lambda_i}{\sum_{i=1}^N \lambda_i} \leq 0.01$$

where  $N$  is the number of dimensions before the PCA and  $\lambda_i$  are the eigenvalues of the covariance matrix.

If all the classes were known a priori, then other methods such as discriminant analysis would be more appropriate for reducing the dimension of the description space. Such an assumption cannot be made in real situations where the classes are defined interactively by the users, so we also avoided making it here, for the groundtruth database we used in our evaluation.

### B. Evaluation of the various selection strategies

In the early stages of a RF session the frontier will usually be very unreliable. Depending on the initialization of the search and on the characteristics of the image classes, the frontier may be much larger than the target class. In such cases, selecting those unlabeled examples that are currently considered by the learner as (potentially) the most relevant can sometimes produce a faster convergence of the frontier during the first few rounds of RF.

<b>MA</b>	<b>Most Ambiguous</b> 1. Select the closest $N$ candidates to the SVM frontier
<b>MAO</b>	<b>Most Ambiguous and Orthogonal</b> 1. Select the set $S$ of the closest $w$ candidates to the SVM frontier 2. Using Eq. 1 choose the $N$ most orthogonal items from $S$
<b>MP</b>	<b>Most Positives</b> 1. Select the most positive $N$ candidates according to the SVM decision function
<b>MPO</b>	<b>Most Positive and Orthogonal</b> 1. Select the set $S$ of the most positive $w$ candidates according to the SVM decision function 2. Using Eq. 1 choose the $N$ most orthogonal items from $S$

TABLE I  
SELECTION STRATEGIES: A COMPARATIVE VIEW.

For this reason, we added to our comparisons the following criteria: select the “most positive” unlabeled examples according to the current decision function of the SVM (denoted as MP criterion) and select the “most positive and orthogonal” unlabeled examples (denoted as MPO). The MPO criterion adds to MP the condition of low redundancy previously described. When comparing the MP criterion to the suggestion in [17] for the early stages of learning, we see that we only focus on the examples for which the values taken by the decision function of the SVM are maximal and completely ignore the examples for which these values are minimal; this is because of the asymmetry of the retrieval context: in general, the number of relevant items is expected to be much lower than the number of irrelevant items. In Table I we present a comparative view of the selection criteria evaluated in this section.

We performed several comparisons between the four selection criteria on the groundtruth database. At every feedback round the emulated user must label as “relevant” or “irrelevant” all the images in a window of size  $ws = 9$ . A search session is initialized by considering one “relevant” example and  $ws - 1$  “irrelevant” examples. Every image in every class serves as the initial “relevant” example for a different RF session, while the associated initial  $ws - 1$  “irrelevant” examples are randomly selected. For reasons that will become apparent later, we use for the comparisons presented here the triangular kernel.

We began by evaluating the different selection criteria on the *first type of scenario* mentioned in the introduction: finding items in a specific target set, by focusing on ranking most of the “relevant” images before the “irrelevant” ones rather than on finding a frontier between the class of interest and the other images. Since the only information available concerns class membership, we do not consider important here the precise ranking of the “relevant” or of the “irrelevant” images.

In order to evaluate the speed of improvement of this ranking, we must use a measure that does not give a prior advantage to one selection criterion. For example, by taking into account already labeled images plus those selected for being labeled during the current round, we should obviously

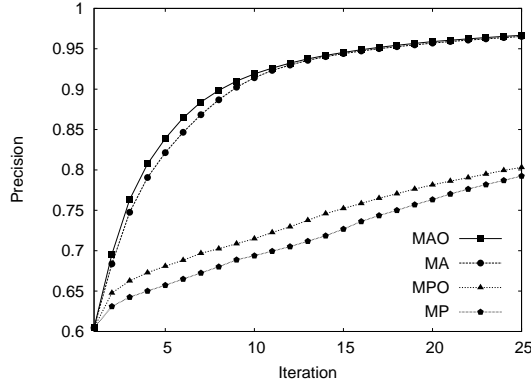


Fig. 4. Comparison of different selection criteria.

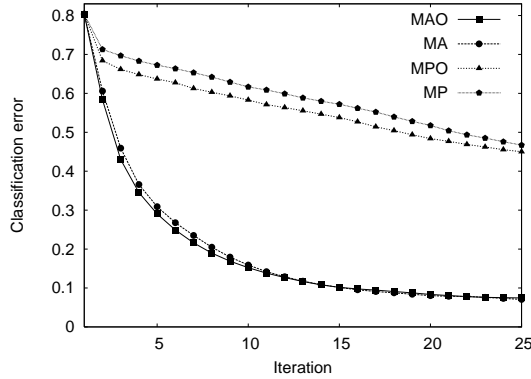


Fig. 5. Evolution of the classification error obtained with different selection criteria.

favor the MP and MPO criteria over MA and MAO. We decided to use instead the following precision measure: at every RF round, we count the number of “relevant” images found in the  $N$  images considered as most positive by the current decision function of the SVM ( $N$  being the number of images in each class).

The evolution of the mean precision during successive RF iterations on the groundtruth database is presented in Fig. 4. The precision value shown is obtained as the mean value over all feedback rounds, each image in the database being used to initiate a new feedback round as described above. Clearly, the reduction of the redundancy between the images selected for labeling improves the results, both for MAO with respect to MA and for MPO with respect to MP. Also, active learning selection criteria (MAO and MA) provide much better results compared to the MP and MPO criteria.

The *second type of scenario* mentioned in the introduction consists of finding a frontier between “relevant” and “irrelevant” images, which can be important for extending the text annotation of some images in the “relevant” class to the others or for mass annotation of images. In this case, we have to evaluate the speed of improvement of the classification. The classification error is defined here as  $p/N + n/N$ , where  $N$  is the class size,  $p$  is the number of false positives and  $n$  is the number of false negatives. In Fig. 5 we can see the evolution of the classification error for the different selection

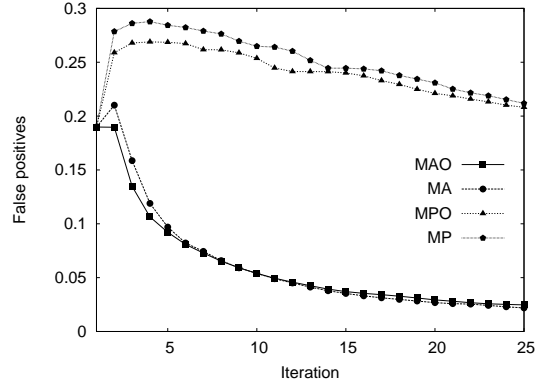


Fig. 6. Evolution of the rate of false positives obtained with different selection criteria.

criteria. As expected, the convergence is fastest for the MAO selection criterion, followed by the MA criterion. Again, active learning selection criteria (MAO and MA) improve greatly the results compared to the MP and MPO criteria. The evolution of the rate of false positive is showed in Fig 6. We see that the convergence is significantly faster for the MAO and MA criteria, which can be explained by the fact that MP and MPO do not focus on the frontier.

These results suggest that the MAO selection criterion should be preferred whenever RF is used for finding a reliable frontier between “relevant” and “irrelevant” images or inter-actively learning new “visual concepts”, but also for the task of ranking “relevant” images before the rest of the database.

As a final remark, we recall that the MAO criterion is as fast as MA while eliminating the redundancy between the selected samples. The overhead imposed over MA is  $O(\text{constant})$ , independent of the size of the database. Since in all our experimental evaluations, MAO performed consistently better than MA at all feedback iterations (although not always by a very large margin, depending on the query target), it can be preferred to the MA criterion because it does not impose any computation penalty.

### C. Insensitivity to scale

In our experiments, for all the kernels we used the L1 norm because experimentally we found it to provide better results than L2. A few other dissimilarity measures (some of which don’t have the properties of a metric) were used in the literature instead of  $\|x_i - x_j\|$ , mainly with the Gaussian kernel and sometimes for variable-length representations of the images. Some of these measures don’t guarantee the convergence of the SVM and we preferred not to use them here.

The sensitivity of the Gaussian and triangular kernel to the scale parameter,  $\gamma$ , parameter with the MP criterion is shown in Fig. 7 (the base of the logarithm is 10). In this figure, the “mean precision” value shown is obtained as the mean over all RF sessions started from each image in the database, as described previously, and over the first 25 feedback iterations. Also, comparisons between these kernels using the MP selection criterion are shown in Fig. 8; for the

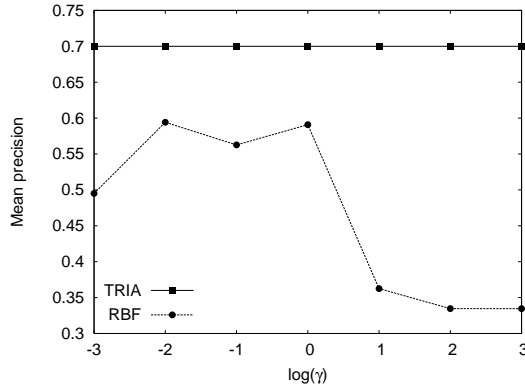


Fig. 7. Sensitivity of the SVM to a scale parameter for the RBF kernel and for the triangular kernel.

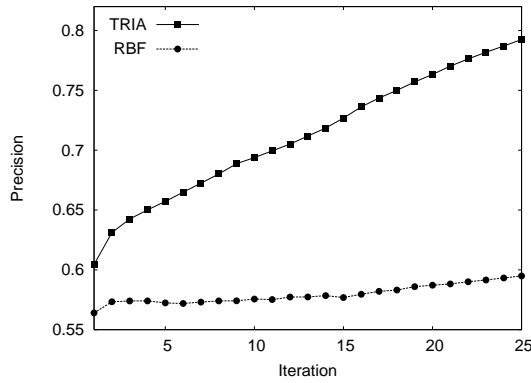


Fig. 8. Comparison of the RBF and triangular kernels with the MP selection criterion.

Gaussian kernels, the scale parameters was set to the overall best value for the database.

From these results we can see that the Gaussian kernel is the one who produces the highest sensitivity to scale for the SVM. Since the classes present in a database often have significantly different scales, any value for the scale parameter will be inadequate for many classes, so the results obtained with this kernel cannot be very good.

Comparatively, the use of the triangular kernel reduces the sensitivity of the SVM to the scale of the data. For real-world applications, the scales of the user-defined classes cannot be known *a priori* and the scale parameter of the kernel cannot be easily adjusted online, thus important variations between classes can be expected for the performance of RF-based retrieval if kernels such as the RBF one are employed. The scale-invariance obtained by the use of the triangular kernel becomes then a highly desirable feature and, as experimental evidence shows, makes this kernel a very good alternative.

#### D. Search examples

To illustrate the effectiveness of our approach for the interactive querying of satellite image databases using SVM-based relevance feedback, we present here some screenshots obtained with our CBIR system<sup>3</sup> [26] on a database of satellite

images.

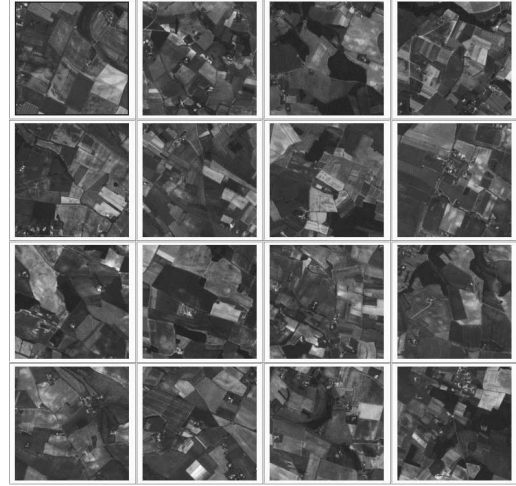


Fig. 9. Query by visual example: looking for cultivated areas in a database of satellite images. These regions are described by patches of uniformly textured areas. In this case, the low level descriptors are sufficiently rich to provide satisfactory answers.

In Fig 9 we present a typical sample of a query by visual example (QBE): searching for “agricultural areas”. The query example is the top-left image and the results are ordered from left to right, top to bottom by decreasing visual similarity to the query image. In this example, the results are promising, but there are situations where QBE does not offer satisfactory answers, as we shall see in the following examples.

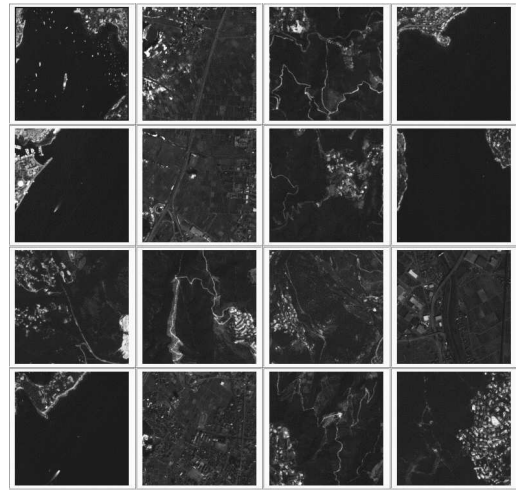


Fig. 10. Query by visual example: looking for urban coast areas. The effects of the semantic gap are clearly visible: the query subject is a mix of thematic classes, difficult to describe directly using low level image signatures.

Fig 10 presents the results of a search for urban coast areas using QBE. The query image is the top left one. As one can see, the results provided by the QBE approach are not very good. Fig 11 presents the results obtained using relevance feedback. When the system returns this screen of results, the user has marked 5 images as “relevant” and 8 images as “irrelevant”. In this case, the system can use the supplementary information provided by the user to estimate the target class

<sup>3</sup><http://www-rocq.inria.fr/imedia/cbir-demo.html>

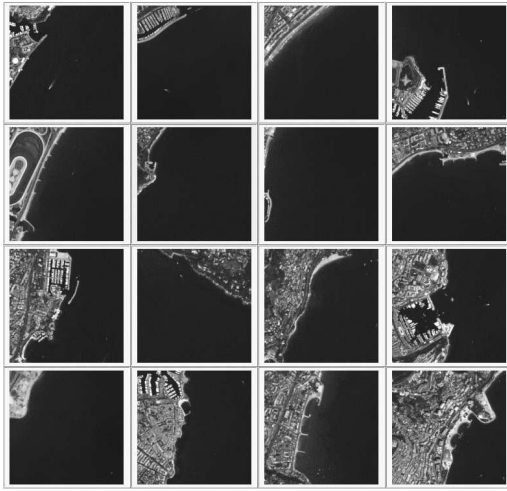


Fig. 11. Search with relevance feedback: looking for urban coast areas (results returned after marking 5 images as “relevant” and 8 images as “irrelevant”). Compared to the QBE case (see Fig 10), the results proposed by relevance feedback are much closer to the target class.

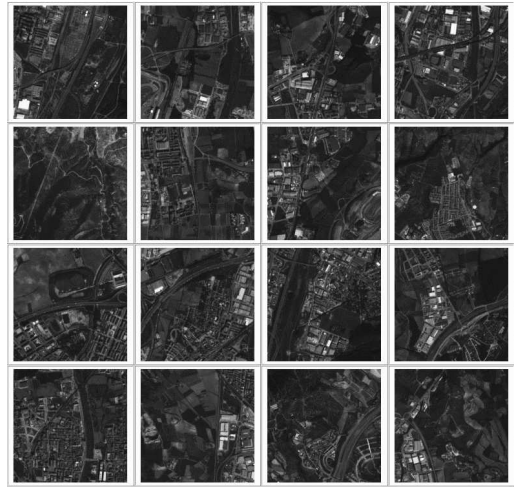


Fig. 13. Search with relevance feedback: looking for urban infrastructures (there are 4 images marked as “relevant” and 6 images as “irrelevant”). Relevance feedback is particularly adapted to satellite images, where the query subject is often complex and semantic in nature.

and the quality of the returned images improves considerably as compared to the QBE case.

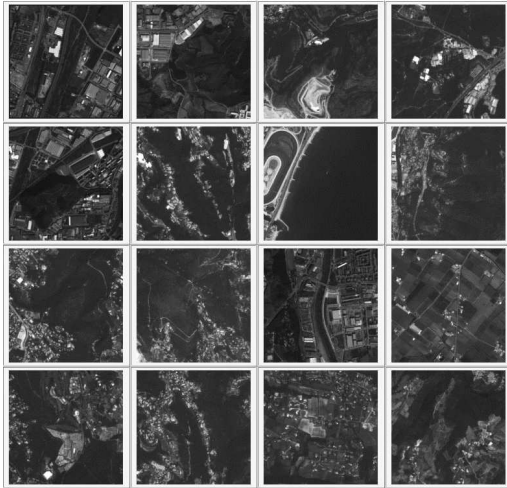


Fig. 12. Query by visual example: looking for urban infrastructures. Due to the complex nature of the query target, the results are strongly affected by the semantic gap.

In the example presented in Fig 12 and Fig 13 the user is searching for urban infrastructures (e.g. highways, bridges, etc.). The QBE approach does not provide satisfactory results, but RF, after positive 4 examples and 6 negative examples, returns a much better selection of images to the user.

## V. CONCLUSIONS AND FURTHER WORK

In this paper we argue that relevance feedback—method successfully used for querying generic image databases—can be employed with very good results with remote sensing databases and we evaluate in this context two improvements of the SVM-based relevance feedback mechanism.

First, to optimize the transfer of information between the user and the system, we put forward an active learning

selection criterion that minimizes redundancy between the candidate images shown to the user. Second, we show that insensitivity of the learning machine to scale of the data is a very desirable property in this context and we show how to obtain it by the use of specific kernel functions. Experimental evaluation of both ranking and classification performance on a groundtruth database of satellite images confirms that our propositions improve the quality of the returned results both in a ranking context and in a classification scenario.

Several interesting directions are available to continue this work. First, in the early stages of the learning, due to the small number of training samples, the frontier may be quite unreliable. Using active learning at this stage may prove sub-optimal. A promising strategy would be to start the relevance feedback session by several MP rounds and then, when the frontier becomes more reliable, continue with MAO. This poses an interesting problem: when to switch from one strategy to another? One idea is to check the evolution of the SVM learner: when the number of support vectors does not change sufficiently from one RF round to the next, the learner is exploring only the inside region delimited by the frontier of the SVM, and thus it may be a good moment to switch to an active learning selection strategy.

Invariance to scale is a very important issue in image retrieval with relevance feedback. Using multiple kernels with different scale parameters, as suggested in [28], is a promising direction to further investigate this subject.

Other directions for improving the relevance feedback mechanism presented here include using relevance feedback to create profiles for different query targets that may be used later by the user to define more complex queries and the evaluation of the RF with real users.

## ACKNOWLEDGEMENTS

The satellite image database used for the evaluation in this paper was kindly provided by Institut Géographique National



France<sup>4</sup>. The authors would like to thank Michel Crucianu, professor at CNAM Paris, for many suggestions and fruitful discussions.

## REFERENCES

- [1] M. Datcu and K. Seidel, "Image information mining: exploration of image content in large archives," in *Aerospace Conference Proceedings*, vol. 3, 2000, pp. 253–264.
- [2] M. Datcu, K. Seidel, and M. Walessa, "Spatial information retrieval from remote-sensing images: information theoretical perspective," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 5, pp. 1431–1445, 2000.
- [3] M. Schroder, H. Rehrauer, K. Siedel, and M. Datcu, "Interactive learning and probabilistic retrieval in remote sensing image archives," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 5, pp. 2288–2298, 2000.
- [4] C. Tusk, K. Koperski, S. Aksoy, and G. Marchisio, "Automated feature selection through relevance feedback," in *IEEE International Geoscience and Remote Sensing Symposium*, 2003, pp. 3691–3693.
- [5] S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, and J. Tilton, "Learning bayesian classifiers for scene classification with a visual grammar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 581–589, 2005.
- [6] A. Parulekar, R. Datta, J. Li, and J. Z. Wang, "Large-scale satellite image browsing using automatic semantic categorization and content-based retrieval," in *Proceedings of the IEEE International Workshop on Semantic Knowledge in Computer Vision, in conjunction with IEEE International Conference on Computer Vision*, October 2005.
- [7] I. E. Alber, M. Farber, N. Yeager, Z. Xiong, , and W. M. Pottenger, "Retrieval of multi-and hyperspectral images using an interactive relevance feedback form of content-based image retrieval," in *In the Proceedings of the 15th Annual International Symposium on Aerospace/Defense Sensing, Simulation, and Controls*, April 2001, pp. 56–66.
- [8] X. S. Zhou and T. S. Huang, "Relevance feedback for image retrieval: a comprehensive review," *Multimedia Systems*, vol. 8, no. 6, pp. 536–544, 2003.
- [9] V. Vapnik, *Estimation of dependences based on empirical data*. Springer Verlag, 1982.
- [10] B. Schölkopf and A. Smola, *Learning with Kernels*. MIT Press, 2002.
- [11] E. Y. Chang, B. Li, G. Wu, and K. Goh, "Statistical learning for effective visual image retrieval," in *Proceedings of the IEEE International Conference on Image Processing (ICIP'03)*, September 2003, pp. 609–612.
- [12] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," in *Proceedings of ICML-00, 17th International Conference on Machine Learning*. Morgan Kaufmann, 2000, pp. 999–1006. [Online]. Available: [citeseer.nj.nec.com/tong00support.html](http://citeseer.nj.nec.com/tong00support.html)
- [13] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proceedings of the 9th ACM international conference on Multimedia*. ACM Press, 2001, pp. 107–118. [Online]. Available: <http://doi.acm.org/10.1145/500141.500159>
- [14] I. J. Cox, M. L. Miller, S. M. Omohundro, and P. N. Yianilos, "An optimized interaction strategy for Bayesian relevance feedback," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 1998, pp. 553–558.
- [15] I. J. Cox, M. L. Miller, T. P. Minka, T. Papathomas, and P. N. Yianilos, "The Bayesian image retrieval system, PicHunter: theory, implementation and psychophysical experiments," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 20–37, January 2000.
- [16] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *Journal of Artificial Intelligence Research*, vol. 4, pp. 129–145, 1996.
- [17] C. Campbell, N. Cristianini, and A. Smola, "Query learning with large margin classifiers," in *Proceedings of ICML-00, 17th International Conference on Machine Learning*. Morgan Kaufmann, 2000, pp. 111–118.
- [18] K. Brinker, "Incorporating diversity in active learning with support vector machines," in *Proceedings of ICML-04, International Conference on Machine Learning*, August 2003, pp. 59–66.
- [19] R. Herbrich, T. Graepel, and C. Campbell, "Bayes point machines," *Journal of Machine Learning Research*, vol. 1, pp. 245–279, 2001. [Online]. Available: [citeseer.nj.nec.com/herbrich01bayes.html](http://citeseer.nj.nec.com/herbrich01bayes.html)
- [20] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support-vector machines for histogram-based image classification," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1055–1064, 1999.
- [21] F. Jing, M. Li, L. Zhang, H.-J. Zhang, and B. Zhang, "Learning in region-based image retrieval," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, 2003.
- [22] C. Berg, J. P. R. Christensen, and P. Ressel, *Harmonic Analysis on Semigroups*. Springer-Verlag, 1984.
- [23] B. Schölkopf, "The kernel trick for distances," in *Advances in Neural Information Processing Systems*, vol. 12. MIT Press, 2000, pp. 301–307.
- [24] F. Fleuret and H. Sahbi, "Scale-invariance of support vector machines based on the triangular kernel," in *3rd International Workshop on Statistical and Computational Theories of Vision*, October 2003.
- [25] C. Vertan and N. Boujemaa, "Upgrading color distributions for image retrieval: can we do better?" in *International Conference on Visual Information Systems (Visual2000)*, November 2000.
- [26] N. Boujemaa, J. Fauqueur, M. Ferecatu, F. Fleuret, V. Gouet, B. L. Saux, and H. Sahbi, "IKONA: Interactive generic and specific image retrieval," in *Proceedings of the International workshop on Multimedia Content-Based Indexing and Retrieval (MMCBIR'2001)*, 2001.
- [27] M. Ferecatu, "Image retrieval with active relevance feedback using both visual and keyword-based descriptors," Ph.D. dissertation, INRIA—Université de Versailles Saint Quentin en Yvelines, France, 2005.
- [28] M. Burl, D. DeCoste, B. Enke, D. Mazzoni, W. Merline, and L. Scharenbroich, "Automated knowledge discovery from simulators," in *SIAM Data Mining Conference*, 2006, pp. 82–93.



**Marin Ferecatu** is a research scientist at INRIA Rocquencourt, IMEDIA research group. He holds a PhD degree in computer science from University of Versailles since 2005. His research focus on active relevance feedback for interactive image retrieval using multiple sources of information, e.g. visual descriptors and text annotations. His scientific interests include Machine Learning, Pattern Recognition, Content Based Image Retrieval, Relevance Feedback, Mining Large Image Databases. He is the author of many papers published in international

journals and conferences.



**Nozha Boujemaa** is a director of research at INRIA (Rocquencourt), and leader of the IMEDIA research group. Her fields of interest are Pattern Recognition, Image Analysis, and Content-Based Image Retrieval. She obtained her PhD in computer science in 1993 and her "Habilitation à Diriger des Recherches" in computer science in 2000. She authored over 70 publications. Dr. Boujemaa has supervised over 30 graduate students at the Master and PhD levels. She took part in various program committees in international conferences (WWW Multimedia, ACM

Multimedia, ICPR, IEEE Fuzzy systems, ICME, RIAO) in the area of visual information retrieval and pattern recognition. She co-chairs the "8th ACM SIGMM MIR" in conjunction with ACM Multimedia 2006 and co-chairs of MMCBIR'01 joint NSF-IBM-BERKLEY-INRIA workshop. She was invited as plenary speaker for four international conferences and organized many special sessions in the topic of machine learning and multimedia retrieval. She is involved in European and national projects covering application areas such as audio-visual archives, photo stocks agency, biodiversity, satellite images and security. She has served as a scientific expert for numerous institutions such as the NSF (ITR), the European Commission (FET), CNRSG (Canada), NWO (Netherlands), A\*Star: Singapore. Nozha Boujemaa serves as expert for the French National Assembly for the LEN law (Loi sur l'Economie Numérique) on illicit internet content filtering.

<sup>4</sup><http://www.ign.fr>