Active learning methods for Interactive Image Retrieval.

Philippe-Henri Gosselin, Matthieu Cord

To cite this version:

Active learning methods for Interactive Image Retrieval

Philippe Henri Gosselin[(1)] and Matthieu Cord[(2)]

(1) gosselin@ensea.fr, ETIS / CNRS, 6, avenue du Ponceau, 95014 Cergy-Pontoise, France
(2) matthieu.cord@lip6.fr, LIP6 / UPMC / CNRS, 104, avenue du President Kennedy, 75016 Paris, France

Abstract—**Active learning methods have been considered with increased interest in the statistical learning community. Initially developed within a classification framework, a lot of extensions are now being proposed to handle multimedia applications. This paper provides algorithms within a statistical framework to extend active learning for online content-based image retrieval (CBIR). The classification framework is presented with experiments to compare several powerful classification techniques in this information retrieval context. Focusing on interactive methods, active learning strategy is then described. The limitations of this approach for CBIR are emphasized before presenting our new active selection process RETIN. First, as any active method is sensitive to the boundary estimation between classes, the RETIN strategy carries out a boundary correction to make the retrieval process more robust. Second, the criterion of generalization error to optimize the active learning selection is modified to better represent the CBIR objective of database ranking. Third, a batch processing of images is proposed. Our strategy leads to a fast and efficient active learning scheme to retrieve sets of online images (query concept). Experiments on large databases show that the RETIN method performs well in comparison to several other active strategies.**

I. INTRODUCTION

Human interactive systems have attracted a lot of research interest in recent years, especially for content-based image retrieval systems. Contrary to the early systems, which focused on fully automatic strategies, recent approaches have introduced human-computer interaction [1], [2]. In this paper, we focus on the retrieval of *concepts* within a large image collection. We assume that a user is looking for a set of images, the *query concept*, within a database. The aim is to build a fast and efficient strategy to retrieve the query concept.

In content-based image retrieval (CBIR), the search may be initiated using a query as an example. The top rank similar images are then presented to the user. Then, the interactive process allows the user to refine his request as much as necessary in a relevance feedback loop. Many kinds of interaction between the user and the system have been proposed [3], but most of the time, user information consists of binary labels indicating whether or not the image belongs to the desired concept. The positive labels indicate *relevant* images for the current concept, and the negative labels *irrelevant* images.

To achieve the relevance feedback process, the first strategy focuses on the query concept updating. The aim of this strategy is to refine the query according to the user labeling. A simple approach, called *query modification*, computes a new query by averaging the feature vectors of relevant images [2]. Another approach, the *query reweighting*, consists in computing a new

similarity function between the query and any picture in the database. A usual heuristic is to weigh the axes of the feature space [4]. In order to perform a better refinement of the similarity function, optimization-based techniques can be used. They are based on a mathematical criterion for computing the reweighting, for instance Bayes error [5], or average quadratic error [6], [7]. Although these techniques are efficient for target search and monomodal concept retrieval, they hardly track complex image concepts.

Performing an estimation of the query concept can be seen as a statistical learning problem, and more precisely as a binary classification task between the relevant and irrelevant classes [8]. In image retrieval, many techniques based on statistical learning have been proposed, as for instance Bayes classification [9], k-Nearest Neighbors [10], Gaussian Mixtures [11], Gaussian random fields [12], or Support Vector Machines [8], [3]. In order to deal with complex and multimodal concepts, we have adopted a statistical learning approach. Additionally, the possibility to work with kernel functions is decisive.

However, a lot of these learning strategies consider the CBIR process as a classical classification problem, without any adaptations to the characteristics of this context. For instance, some discriminative classifiers exclusively return binary labels when real values are necessary for CBIR ranking purposes. Furthermore, the system has to handle classification with few training data, especially at the beginning of the search, where the query concept has to be estimated in a database of thousands of images with only a few examples. Active learning strategies have been proposed to handle this type of problem. Another point concerns the class sizes, since the query concept is often a small subset of the database. In contrast to more classical classification problems, relevant and irrelevant classes are highly imbalanced (up to factor 100). Depending on the application context, computational time has also to be carefully considered when an online retrieval algorithm is designed. To address this problem, we assume that any learning task must be at most $O(n)$, where n is the size of the database.

In this paper, we focus on statistical learning techniques for interactive image retrieval. We propose a scheme to embed different active learning strategies into a general formulation. The originality of our approach is based on the association of 3 components:
 • Boundary correction, which corrects the noisy classification boundary in the first iterations;
 • Average Precision maximization, which selects the im-

ages so that classification and Mean Average Precision are enhanced;

• Batch selection, which addresses the problem of the selection of multiple images in the same feedback iteration.

We also propose a pre-selection technique to speed up the selection process, which leads to a computational complexity negligible compared to the size of the database for the whole active learning process. All these components are integrated in our retrieval system, called RETIN.

In this scope, we first present the binary classification and kernel framework to represent complex class distributions (section II). Powerful classification methods and well-defined kernels for global image signatures are evaluated on real database experiments. Secondly, we present active learning to interact with the user (section III), then introduce the RETIN active learning scheme in section IV, where its components are described in sections V, VI and VII. Finally, we compare our method to existing ones using real scenario on large databases (section VIII).

II. CLASSIFICATION FRAMEWORK

In the classification framework for CBIR, retrieving image concepts is modeled as a two-class problem: the relevant class, the set of images in the searched concept, and the irrelevant class, composed by the remaining database.

Let $\{\mathbf{x}_i\}_{1,n}$ be the n image indexes of the database. A training set is expressed from any user label retrieval session as $\mathcal{A}_\mathbf{y} = \{(\mathbf{x}_i, y_i)_{i=1,n} \mid y_i \neq 0\}$, where $y_i = 1$ if the image $\mathbf{x}_i$ is labeled as relevant, $y_i = -1$ if the image $\mathbf{x}_i$ is labeled as irrelevant (otherwise $y_i = 0$). The classifier is then trained using these labels, and a relevance function $f_{\mathcal{A}_\mathbf{y}}(\mathbf{x}_i)$ is determined in order to be able to rank the whole database.

Image indexes contain a summary of visual features, in this paper histograms of colors and textures.

A. Classification methods for CBIR

Bayes and probabilistic classifiers are the most used classification methods for CBIR [13], [14], [15]. They are interesting since they are directly "relevance oriented", *i.e.* no modifications are required to get the probability of an image to be in the concept. However, since Gaussian-based models are generally used, they focus on estimating the center of each class, and then are less accurate near the boundary than discriminative methods, which is an important aspect for active learning. Furthermore, because of the unbalance of training data, the tuning of the irrelevant class is not trivial.

k-Nearest Neighbors determines the class of an image considering its nearest neighbors. Despite its simplicity, it still has real interest in terms of efficiency, especially for the processing of huge databases [16].

Support Vector Machines are classifiers which has known much success in many learning tasks during the past years. They were introduced with kernel functions in the statistical community [17], and were quickly used by researchers of the CBIR community [18]. SVMs select the optimal separating hyperplane which has the largest margin and hence the lowest vicinal risk. Their high ability to be used with a kernel function provides many advantages that we describe in the next section.

Fisher Discriminant Analysis also uses hyperplane classifiers, but aims at minimizing both the size of the classes and the inverse distance between the classes [19]. Let us note that in experiments, they give results similar to SVMs [20].

B. Feature distribution and kernel framework

In order to compute the relevance function $f_{\mathcal{A}_\mathbf{y}}(\mathbf{x}_i)$ for any image $\mathbf{x}_i$, a classification method has to estimate the density of each class and/or the boundary. This task is mainly dependent on the shape of the data distribution in the feature space. In the CBIR context, relevant images may be distributed in a single mode for one concept, and in a large number of modes for another concept, thereby inducing nonlinear classification problems.

Gaussian Mixtures are highly used in the CBIR context, since they have the ability to represent these complex distributions [21]. However, in order to get an optimal estimation of the density of a concept, data have to be distributed as Gaussian (limitation for real data processing). Furthermore, the large number of parameters required for Gaussian Mixtures leads to high computational complexity. Another approach consists in using the kernel function framework [22]. These functions were introduced in the statistical learning community with the Support Vector Machines. The first objective of this framework is to map image indexes $\mathbf{x}_i$ to vectors $\Phi(\mathbf{x}_i)$ in a Hilbert space, and hence it turns the non-linear problem into linear ones. It is also different from other linearization techniques by its ability to implicitly work on the vectors of the Hilbert space. Kernel methods never explicitly compute the vectors $\Phi(\mathbf{x}_i)$, but only work on their dot product $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$, and hence allow to work on very large or infinite Hilbert spaces. Furthermore, learning tasks and distribution of data can be separated. Under the assumption that there is a kernel function which maps our data into a linear space, all linear learning techniques can be used on any image database.

We denote by $k(\mathbf{x}_i, \mathbf{x}_j)$ the value of the kernel function between images $\mathbf{x}_i$ and $\mathbf{x}_j$. This function will be considered as the default similarity function in the following.

C. Comparison of classifiers

We have experimented several classification methods for CBIR: Bayes classification [9] with a Parzen density estimation, k-Nearest Neighbors [10], Support Vector Machines [8], Kernel Fisher Discriminant [19], and also a query-reweighting strategy [7]. Databases, scenario, evaluation protocol and quality measurement are detailed in the appendix. The results in terms of Mean Average Precision are shown on Fig. II-C according to the training set size (we omit the KFD which gives results very close to inductive SVMs) for both ANN and Corel databases.

One can see that the classification-based methods give the best results, showing the power of statistical methods over geometrical approaches, like the one reported here (similarity refinement method). The SVM technique performs slightly
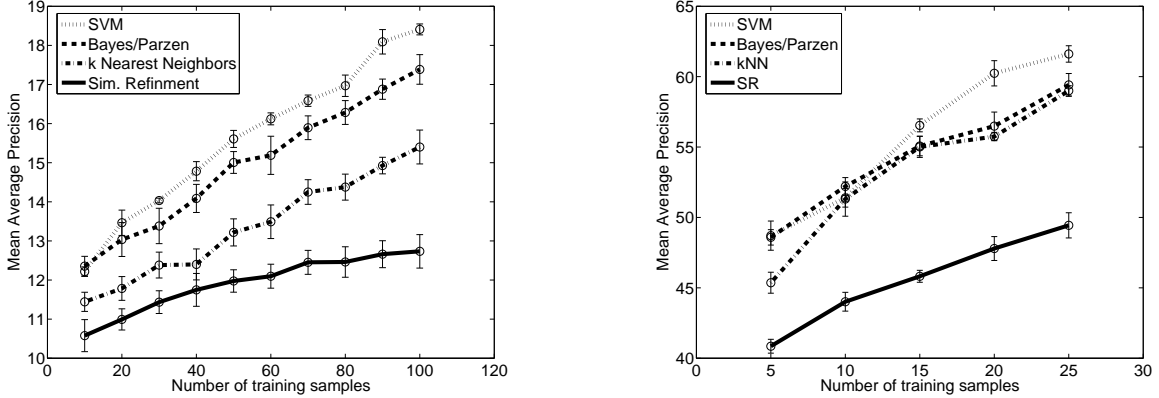
Fig. 1.   Mean Average Precision(%) for a classification from randomly selected examples on the Corel (left) and the ANN (right) photo database.

better than others in this context. As they have a strong mathematical framework and efficient algorithmic implementation, SVMs are used as the default method in the RETIN system.

We also made some comparisons between different kernels (linear, Polynomial, Triangle, Gaussian $L1$, Gaussian $L2$ and Gaussian χ^2). In our experiments, the Gaussian χ^2 gives the best performance, which is not a surprising result since histograms are used as image signatures, and the χ^2 distance is dedicated for comparing distributions. In the following experiments, we always use a Gaussian kernel with a χ^2 distance.

Remark: As the whole data set is available during the training, it could be interesting to consider a semi-supervised or transductive framework. For instance, there are extended versions of Gaussian Mixtures [11], [13] and transductive SVM [23]. We have experimented with these methods. The computational time is very high and no improvement has been observed [24], [25]. Transduction does not seem to be an interesting approach for CBIR as already observed in [3].

Anyway, the global performances remain very low for any proposed methods for the Corel experiments. The MAP is under 20%, even when the number of training data is up to 100 for Corel. The Corel database is much bigger than ANN and the simulated concepts are more difficult. The training set remains too small to allow classifiers to efficiently learn the query concept. Active learning is now considered to overcome this problem.

III. ACTIVE LEARNING

Active learning is close to supervised learning, except that training data are not independent and identically-distributed variables. Some of them are added to the training set thanks to a dedicated process. In this paper, we only consider selective sampling approaches from the general active learning framework. In this context, the main challenge is to find data that, once added to the training set, will allow us to achieve the best classification function. These methods have been introduced to perform good classifications with few training data in comparison to the standard supervised scheme.

When the learner can only choose new data in a pool of unlabeled data, it is called *pool-based* active learning

framework [26]. In CBIR, the whole set of images is available anytime. These data will be considered as the pool of unlabeled data during the selective sampling process.

A. Example of active strategies

Fig. 2 shows the interest of a selection step. In this example, the images are represented by 2D feature vectors, the white circles are images the user is looking for, and the black circles are the images the user is not interested in. At the beginning, the user provided two labels, represented in figures by larger circles (*cf.* Fig 2(a)). These two labels allow the system to compute a first classification. In classical relevance feedback systems, a common way of selection was to label the most relevant pictures returned by the system. As one can see in Fig 2(b), this choice is not effective, since in that case the classification is not improved. Other types of selection of new examples may be considered. For instance, in Fig 2(c), the active learning selection working on uncertainty is proposed: the user labels the pictures the closest to the boundary, resulting in an enhanced classification in that case Fig 2(c).

B. Optimization scheme

Starting from the image signatures $\{\mathbf{x}_i\}_{1,n}$, the training set $\mathcal{A}_{\mathbf{y}} = \{(\mathbf{x}_i, y_i)_{i=1,n} \mid y_i \neq 0\}$ and the relevance function $f_{\mathcal{A}_{\mathbf{y}}}$ defined in section II, new notations are introduced for the teacher $s : X \to \{-1, 1\}$ that labels images as -1 or 1, the indexes of the labeled images I, and the unlabeled ones $\bar{I}$.

The active learning aims at selecting the unlabeled data $\mathbf{x}_{i^\star}$ that will enhance the most the relevance function f trained with the label $s(\mathbf{x}_{i^\star})$ added to $\mathcal{A}_{\mathbf{y}}$. To formalize this selection process as a minimization problem, a cost function $g_{\mathcal{A}_{\mathbf{y}}}$ is introduced. According to any active learning method, the selected image is $\mathbf{x}_{i^\star}$ minimizing $g_{\mathcal{A}_{\mathbf{y}}}(\mathbf{x})$ over the pool of unlabeled images:

$$i^\star = \underset{i \in \bar{I}}{\operatorname{argmin}} \left(g_{\mathcal{A}_{\mathbf{y}}}(\mathbf{x}_i) \right) \qquad (1)$$

(a) Initial training set.   (b) Label the most relevant data.   (c) Label the most uncertain data.
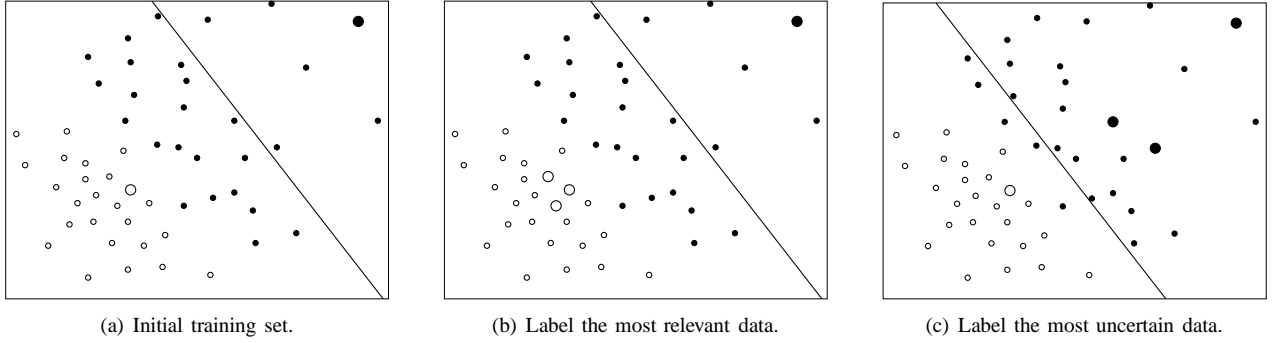
Fig. 2.   Active learning illustration. A linear classifier is computed for white (relevant) and black (irrelevant) data classification. Only the large circles are used for training, the small ones represent unlabeled data. The line is the boundary between classes after training. (a) represents the initial boundary; In (b) two new data (the closest to the relevant one) have been added to the training set; the boundary is unchanged. In (c), the most uncertain data (closest to the boundary) are added to the training data; the boundary significantly moved, and provides a better separation between black and white data.

C. Active Learning Methods

Two different strategies are usually considered for active learning: the uncertainty-based sampling, that selects the images for which the relevance function is the most uncertain (Fig 2), and the error reduction strategy, that aims at minimizing the generalization error of the classifier.

According to this distinction, we have selected two popular active learning strategies for presentation and comparison.

1) Uncertainty-based sampling: In our context of binary classification, the learner of the relevance function has to classify data as relevant or irrelevant. Any data in the pool of unlabeled samples may be evaluated by the learner. Some are definitively relevant, others irrelevant, but some may be more difficult to classify. Uncertainty-based sampling strategy aims at selecting unlabeled samples that the learner is the most uncertain about.

To achieve this strategy, a first approach proposed by Cohn [27] uses several classifiers with the same training set, and selects samples whose classifications are the most contradictory. Another solution consists in computing a probabilistic output for each sample, and selecting the unlabeled samples with the probabilities closest to 0.5 [28]. Similar strategies have also been proposed with SVM classifier [29], with a theoretical justification [18], and with nearest neighbor classifier [30].

In any case, a relevance function $f_{\mathcal{A}_\mathbf{y}}$ is trained. This function may be adapted from a distribution, a membership to a class (distance to the hyperplane for SVM), or a utility function. Using this relevance function, uncertain data $\mathbf{x}$ will be close to 0: $f_{\mathcal{A}_\mathbf{y}}(\mathbf{x}) \sim 0$.

The solution to the minimization problem in eq. 1 is:

$$i^\star = \operatorname*{argmin}_{i \in \bar{I}} \left(|f_{\mathcal{A}_\mathbf{y}}(\mathbf{x}_i)| \right) \qquad (2)$$

The efficiency of these methods depends on the accuracy of the relevance function estimation close to the boundary between relevant and irrelevant classes.

2) Error Reduction-based strategy: active learning strategies based on error reduction [26] aim at selecting the sample that, once added to the training set, minimizes the error of generalization of the new classifier.

Let denote $P(y|\mathbf{x})$ the (unknown) probability of sample $\mathbf{x}$ to be in class y (relevant or irrelevant), and $P(\mathbf{x})$ the (also

unknown) distribution of the images. With $\mathcal{A}_\mathbf{y}$, the training provides the estimation $\hat{P}_{\mathcal{A}_\mathbf{y}}(y|\mathbf{x})$ of $P(y|\mathbf{x})$, and the expected error of generalization is:

$$E(\hat{P}_{\mathcal{A}_\mathbf{y}}) = \int_\mathbf{x} L(P(y|\mathbf{x}), \hat{P}_{\mathcal{A}_\mathbf{y}}(y|\mathbf{x})) dP(x)$$

with L a loss function which evaluates the loss between the estimation $\hat{P}_{\mathcal{A}_\mathbf{y}}(y|\mathbf{x})$ and the true distribution $P(y|\mathbf{x})$.

The optimal pair $(\mathbf{x}_i^\star, y_i^\star)$ minimizes this expectation over the pool of unlabeled images:

$$(\mathbf{x}_i^\star, y_i^\star) = \operatorname*{argmin}_{(\mathbf{x}_i, y_i), i \in \bar{I}} \left(E(\hat{P}_{\mathcal{A}_{\mathbf{y}+(\mathbf{x}_i, y_i)}}) \right) \qquad (3)$$

and $\mathcal{A}_\mathbf{y}^\star = \mathcal{A}_{\mathbf{y}+(\mathbf{x}_i^\star, y_i^\star)}$.

As the expected error $E()$ is not accessible, the integral over $P(\mathbf{x})$ is usually approximated using the unlabeled set. Roy and McCallum [26] also propose to estimate the probability $P(y|\mathbf{x})$ with the relevance function provided by the current classifier. With a $0/1$ loss function L, the estimation of the expectation is expressed for any $\mathcal{A}$:

$$\hat{E}(\hat{P}_\mathcal{A}) = \frac{1}{|\bar{I}|} \sum_{\mathbf{x}_i, i \in \bar{I}} \left(1 - \max_{y \in \{-1,1\}} \hat{P}_\mathcal{A}(y|\mathbf{x}_i) \right)$$

Furthermore, as the labels $s(\mathbf{x}_i)$ on $\bar{I}$ are unknown, they are estimated by computing the expectation for each possible label. Hence, the cost function g is given:

$$g_{\mathcal{A}_\mathbf{y}}(\mathbf{x}) = \sum_{y \in \{-1,1\}} \hat{E}(\hat{P}_{\mathcal{A}_\mathbf{y}+(\mathbf{x},y)}) \hat{P}_\mathcal{A}(y|\mathbf{x}) \qquad (4)$$

The following relation between $\hat{P}_{\mathcal{A}_\mathbf{y}}(y|\mathbf{x})$ and $f_{\mathcal{A}_\mathbf{y}}(\mathbf{x})$ is used:

$$\hat{P}_{\mathcal{A}_\mathbf{y}}(y|\mathbf{x}) = \frac{y}{2}(f_{\mathcal{A}_\mathbf{y}}(\mathbf{x}) + y)$$

D. Active Learning in CBIR context

Active learning methods are generally used for classification problems where the training set is large and classes are well balanced. However, our context is somewhat different :
- Unbalance of classes: the class of relevant images (the searched concept), is generally 20 to 100 times smaller than the class of irrelevant images. As a result, the boundary is
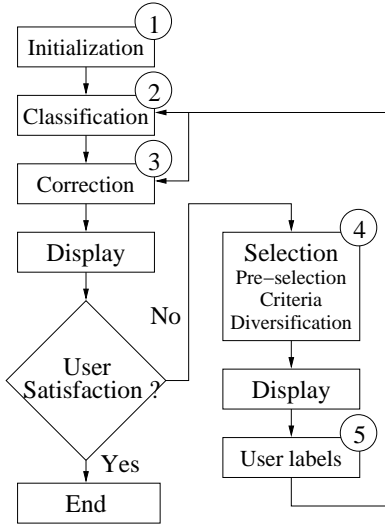
Fig. 3. RETIN active learning scheme.

very inaccurate, especially in the first iterations of relevance feedback, where the size of the training set is dramatically small. Many methods become inefficient in this context, and the selection is then somewhat random.

• Selection criterion: whenever minimizing the error of classification is interesting for CBIR, this criterion does not completely reflect the user satisfaction. Other utility criteria closer to this, such as Precision, should provide more efficient selections.

• Batch selection: we have to propose more than one image to label between two feedback steps, contrary to many active learning techniques which are only able to select a single image.

The computation time is also an important criterion for CBIR in generalized applications, since people will not wait several minutes between two feedback steps. Furthermore, a fast selection allows the user to provide more labels in the same time. Thus, it is more interesting to use a less efficient but fast method than a more efficient but highly-computational one.

To take into account these characteristics, we propose in the next section an active learning strategy dedicated to CBIR context, based on the optimization of function $g_{\mathcal{A}_\mathbf{y}}$, as proposed in eq. 2 or eq. 4. Aside from the active learning context, some other learning approaches have been proposed in Bayesian framework. Dedicated to target search (where the user is looking for a single image) a probability for each image to be the target is updated thanks to user interaction [31], [32].

IV. RETIN ACTIVE LEARNING SCHEME

We propose an active learning scheme based on binary classification in order to interact with a user looking for image concepts in databases. The scheme in summarized in figure 3.

1) Initialization: A retrieval session is initialized from one image brought by the user. The features are computed on that new image and added to the database. This image is then labeled as relevant, and the closest pictures are shown to the user. Note that other initializations could be used, for instance with keywords.

2) Classification: A binary classifier is trained with the labels the user gives. We use a SVM with a Gaussian χ^2 kernel, since it has revealed being the most efficient [33], [34]. The result is a function $f_{\mathcal{A}_\mathbf{y}}(\mathbf{x}_i)$ which returns the relevance of each image $\mathbf{x}_i$, according to the examples $\mathcal{A}_\mathbf{y}$.

3) Boundary Correction: We add an active correction to the boundary in order to deal with the few training data and the imbalance of the classes. We present this technique in the next section.

4) Selection: In the case where the user is not satisfied with the current classification, the system selects a set of images the user should label. The selection must be such so that the labeling of those images provides the best performances. We divide the selection into three steps.

The first step aims at reducing the computational time, by pre-selecting some hundreds of pictures $\mathbf{x}_i$, $i \in J$ which may be in the optimal selection set. We propose to pre-select the closest pictures to the (corrected) boundary. This process is computed very fast, and the uncertainty-based selection method have proved its interest in CBIR context.

The second step is the computation of the selection criterion $g_{\mathcal{A}_\mathbf{y}}(\mathbf{x}_i)$ for each pre-selected image. We use an uncertainty-based selection criterion which aims at maximizing the Average Precision, whose details are presented in section VI.

The third step computes the batch selection. The method presented in section VII selects q images $\mathbf{x}_i$, $i \in I^\star$ using the previously computed criteria $g_{\mathcal{A}_\mathbf{y}}(\mathbf{x}_i)$.

5) Feedback: The user labels the selected images, and a new classification and correction can be computed. The process is repeated as many times as necessary.

V. ACTIVE BOUNDARY CORRECTION

During the first steps of relevance feedback, classifiers are trained with very few data, about 0.1% of the database size. At this stage, classifiers are not even able to perform a good estimation of the size of the concept. Their natural behavior in this case is to divide the database into two parts of almost the same size. Each new sample changes the estimated class of hundreds, sometimes thousands of images. Selection is then close to a random selection.

A solution is to ensure that the retrieval session is initialized with a minimum of examples. For instance, Tong proposes to initialize with 20 examples [33]. Beyond the question of the minimum number of examples required for a good starting boundary, initial examples are not always available without some third-party knowledge (for instance, keywords).

Thus, we propose a method to correct the boundary in order to reduce this problem. A first version of this method was proposed in [35], but we introduce here a generalization of this one. The correction is a shift of the boundary:

$$\hat{f}_{\mathcal{A}_{\mathbf{y}_t}}(\mathbf{x}_i) = f_{\mathcal{A}_{\mathbf{y}_t}}(\mathbf{x}_i) - b_t \tag{5}$$

with $f_{\mathcal{A}_{\mathbf{y}_t}}$ the relevance function, and b_t the correction at feedback step t.

We compute the correction to move the boundary towards the most uncertain area of the database. In other words,

we want a positive value of $\hat{f}_{\mathcal{A}_{\mathbf{y}_t}}()$ for any image in the concept, and a negative value for others. In order to get this behavior, a ranking of the database is computed: $O_{\mathcal{A}_{\mathbf{y}_t}} = \text{argsort}(f_{\mathcal{A}_{\mathbf{y}_t}}(X))$ with $\text{argsort}(\mathbf{v})$ a function returning the indexes of the sorted values of $\mathbf{v}$.

Let r_t be the rank of the image whose relevance is the most uncertain. We have:

$$\underbrace{\mathbf{x}_{O_1}, \mathbf{x}_{O_2},}_{\text{Concept center}} \ldots, \underbrace{\mathbf{x}_{O_{r_t-1}}, \mathbf{x}_{O_{r_t}}, \mathbf{x}_{O_{r_t+1}},}_{\text{Zone of uncertainty}} \ldots, \underbrace{\mathbf{x}_{O_{n-1}}, \mathbf{x}_{O_n}}_{\text{Less relevant images}}$$

Our correction is expressed by: $b_t = f(\mathbf{x}_{O_{r_t}})$.

In order to compute the correction, we propose an algorithm based on an adaptive tuning of r_t during the feedback steps. The value at the $(t+1)$th iteration is computed considering the set of labels provided by the user at the current iteration t.

Actually, we suppose that the best threshold corresponds to the searched boundary. Such a threshold allows to present as many relevant images as irrelevant ones. Thus, if and only if the set of the selected images is well balanced (between relevant and irrelevant images), then the threshold r_t is good. We exploit this property to tune r_t.

At the tth feedback step, the system is able to classify images using the current training set. The user gives new labels for images $\mathbf{x}_{O_{r_t-q/2}}, \mathbf{x}_{O_{r_t}}, \mathbf{x}_{O_{r_t+q/2}}$, and they are compared to the current classification. If the user mostly gives relevant labels, the system should propose new images for labeling around a higher rank to get more irrelevant labels. On the contrary, if the user mostly gives irrelevant labels, thus classification does not seem to be good to rank r_t, and new images for labeling should be selected around a lower rank (to get more relevant labels). In order to get this behavior, we introduce the following update rule:

$$r_{t+1} = r_t + h(\hat{f}_{\mathcal{A}_{\mathbf{y}_t}}, I_t^{\star}, \mathbf{y}_{t+1}) \tag{6}$$

with $I_t^{\star}$ the set of indexes of selected images at step t and function $h()$ such as:

- If we have as many positive labels as negative ones in $I_t^{\star}$, no changes are needed since the boundary is certainly well placed. Thus, $h()$ should be close to zero.
- If we have much more positive than negative labels, the boundary is certainly close to the center of the relevant class. In order to find more negative labels, a better boundary should correspond to a further rank r_{t+1}, which means that $h()$ should be positive.
- On the contrary, if we have many negative labels, the boundary is certainly too far from the center of the relevant class. In this case, $h()$ should be negative.

The easiest way to carry out a function h with such a behavior is to compute the difference between the number of positive and negative labels $y_{i,t+1}$, $i \in I_t^{\star}$. However, it is interesting to take into account the error between the current relevance $\hat{f}_{\mathcal{A}_{\mathbf{y}_t}}(\mathbf{x}_i)$ of a new labeled image $\mathbf{x}_i \in I_t^{\star}$ and the label $y_{i,t+1}$ provided by the user. Indeed, if an image is labeled as positive ($y_i = 1$) and its current relevance is close to 1, this label should not be taken into account. On the contrary, if an image is labeled as negative and its current relevance is close to 1, it

means that the current boundary is too far from the center of the relevant class. In this case a high change of r_t is required.

This behavior for h can be obtained by the computation of the average error between the current relevance $\hat{f}_{\mathcal{A}_{\mathbf{y}_t}}(\mathbf{x}_i)$ of a new labeled image $i \in I_t^{\star}$ and the label $y_{i,t+1}$ provided by the user:

$$h(\hat{f}_{\mathcal{A}_{\mathbf{y}_t}}, I_t^{\star}, \mathbf{y}_{t+1}) = \sum_{i \in I_t^{\star}} \left(y_{i,t+1} - \hat{f}_{\mathcal{A}_{\mathbf{y}_t}}(\mathbf{x}_i) \right) \tag{7}$$

Remark: when using SVMs for computing the classifier $f_{\mathcal{A}_{\mathbf{y}_t}}$, this correction process is close to the computation of parameter b in the SVM decision function :

$$f_{\mathcal{A}_{\mathbf{y}_t}}^{SVM}(\mathbf{x}_i) = \sum_j \alpha_j y_j k(\mathbf{x}_j, \mathbf{x}_i) + b \tag{8}$$

Parameter b can be computed using the KKT conditions [20], for instance if $\mathbf{x}_i$ is a support vector : $b = y_i - f_{\mathcal{A}_{\mathbf{y}_t}}^{SVM}(\mathbf{x}_i)$.

Hence, one can see that our boundary correction is close to the computation of parameter b of the SVM decision function (Eq. 8), except that we are considering vectors $\mathbf{x}_i$ out of the training set ($i \in I_t^{\star}$).

VI. AVERAGE PRECISION MAXIMIZATION

Any active learning method aims at selecting samples which decreases the error of classification. In CBIR, users are interested in database ranking. A usual metric to evaluate this ranking is the Average Precision (see appendix for definition).

A. Average Precision vs Classification Error

We made experiments to evaluate the difference between the direct optimization of the Average Precision and the classification error.

For the minimization of the classification error scheme, the optimal pair $(\mathbf{x}_i^{\star}, y_i^{\star})$ is the one which satisfies the following equation over the unlabeled data pool $\bar{I}$ (*cf.* Eq. 3):

$$(\mathbf{x}_i^{\star}, y_i^{\star}) = \operatorname*{argmin}_{(\mathbf{x}_i, y_i), i \in \bar{I}} \left(E(\hat{P}_{\mathcal{A}_{\mathbf{y}+(\mathbf{x}_i, y_i)}}) \right)$$

where $E(\hat{P}_{\mathcal{A}_{\mathbf{y}+(\mathbf{x}_i, y_i)}})$ is the expected error of generalization of the classifier trained with set $\mathcal{A}_{\mathbf{y}+(\mathbf{x}_i, y_i)}$.

In the case of the maximization of the Average Precision, the expression is close to the previous one, except that we consider the Average Precision $M_{\hat{O}_{\mathcal{A}_{\mathbf{y}}}}$:

$$(\mathbf{x}_i^{\star}, y_i^{\star}) = \operatorname*{argmax}_{(\mathbf{x}_i, y_i), i \in \bar{I}} \left(M_{\hat{O}_{\mathcal{A}_{\mathbf{y}+(\mathbf{x}_i, y_i)}}} \right)$$

where $\hat{O}_{\mathcal{A}_{\mathbf{y}+(\mathbf{x}_i, y_i)}}$ is the ranking of the database computed from the classifier output trained with the set $\mathcal{A}_{\mathbf{y}+(\mathbf{x}_i, y_i)}$.

We compared these two criteria on a reference database where all the labels are known. One can notice that these criteria cannot be used in real applications, where most of the labels are unknown, and are only used here for evaluation purpose :

Training set size	25	50	75	100
Ave. Precision Maximization	50%	58%	62%	65%
Classif. Error Minimization	31%	38%	42%	45%

First of all, one can see that for both criteria, the MAP always increases, which means that an iterative selection of examples will always increase the performance of the system. Furthermore, whenever the classification error method increases the MAP, the technique maximizing the Average Precision performs significantly better with a gain around 20%. This large difference with the error minimization is a motivation to develop Average Precision-based strategies even if the true MAP is not available and has to be approximated. The aim here is to select the images that will maximize the Average Precision, and hence estimate $M_{\hat{O}_{\mathcal{A}_\mathbf{y}}}$.

However, the estimation of $M_{\hat{O}_{\mathcal{A}_\mathbf{y}}}$ is particularly difficult with the kind of samples we have chosen. Indeed, we opted for binary labels, which are simple enough for any non-expert user. Those labels are well adapted for classification, but they are less suitable with the estimation of Average Precision: this criterion is based on a ranking of the database, but binary labels do not give any information in such a way.

B. Precision-Oriented Selection

As explained before, the straight estimation of the Average Precision is not efficient in our context. We opted for a different strategy but have kept in mind the objective of optimization of the ranking. We experimentally noticed that selecting images close to the boundary between relevant and irrelevant classes is definitively efficient, but when considering a subset of images close to the boundary, the active sampling strategy of selecting the closest image inside this subset is not necessarily the best strategy. We propose to consider the subest of images close to the boundary, and then to use a criterion related to the Average Precision in order to select the winner image of the selective sampling strategy.

In order to compute a score related to Average Precision, we propose to consider the sub-database of the labeled pictures. We have the ground truth for this sub-database, since we have the labels of all its images. Thus, it becomes feasible to compute the Average Precision on this sub-database, without any estimation. We still need a ranking of this sub-database in order to compute the Average Precision. We compute the similarity $k(\mathbf{x}_i, \mathbf{x})$ of an unlabeled image $\mathbf{x}_i$ to any labeled images $\mathbf{x}$ (using the kernel function k as the similarity function), and then rank the labeled images according to these similarity values. The resulting factor $m_{\mathcal{A}_\mathbf{y}}(\mathbf{x}_i)$ is the Average Precision on the sub-database with this ranking The aim of this factor $m_{\mathcal{A}_\mathbf{y}}(\mathbf{x}_i)$ is to support the picture $\mathbf{x}_i$ that, once labeled, will have the most chance to increase the Average Precision.

Our final cost function achieves a tradeoff between getting the closest image and optimizing the Average Precision:

$$g_{\mathcal{A}_\mathbf{y}}(\mathbf{x}_i) = |\hat{f}_{\mathcal{A}_\mathbf{y}}(\mathbf{x}_i)| \times (1 - m_{\mathcal{A}_\mathbf{y}}(\mathbf{x}_i)) \qquad (9)$$

Then, using this factor, in the case where the user labels the selected image as positive, the image and its neightbors will be well ranked. Since we selected the image which, at the same time, is the closest to positive labels and furthest from negative labels, the new labeled image (and its neightbors) will be well ranked, and will probably not bring irrelevant images in the top of the ranking.

Let us note that the idea of combining "pessimist" (like uncertainty-based) and "optimist" (like our precision oriented factor) strategies seems to be an interesting way of investigation. We have also tested a combination of uncertainty-based and error reduction by selecting images with the algorithm of Roy & McCallum[26] but only testing on the 100 closest images to the boundary. This method works a little better than the two other ones (in terms of MAP), but the difference is not as large as when combining with a precision-oriented factor.

VII. BATCH SELECTION

The aim of batch selection is to select the set that minimizes the selection criterion over all the sets of q unlabeled images. As for the selection of one image, this minimization has to be estimated. Because of the constraint in terms of complexity of the CBIR context, a direct estimation cannot be performed. A naive extension of the selection for more than one sample is to select the q samples that minimize function g. However, a better batch selection can be achieved by selecting samples in an iterative way [36]. Although this strategy is sub-optimal, it requires little computation in comparison to an exhaustive search of the best subset.

Actually, our algorithm selects q pictures one by one, and prevents from the selection of close samples:

$$
\begin{aligned}
&I^\star = \{\} \\
&\textbf{for } l \in [1..q] \\
&\quad i^\star = \underset{i \in J - I^\star}{\mathrm{argmin}} \left(g_{\mathcal{A}_\mathbf{y}}(\mathbf{x}_i) + \max_{j \in I \cup I^\star} k(\mathbf{x}_i, \mathbf{x}_j) \right) \\
&\quad I^\star = I^\star \cup \{i^\star\} \\
&\textbf{endfor}
\end{aligned}
$$

with $k(\mathbf{x}_i, \mathbf{x}_j)$ kernel function between sample $\mathbf{x}_i$ and sample $\mathbf{x}_j$.

The efficiency of this batch selection is the result of the efficiency of current classification techniques, where the labeling of one or more close samples provides almost the same classification output. Hence, this batch selection process provides more diversity in the training set.

Our batch selection does not depend on a particular technique of selection of only one image, hence it may be used with all the previously described active learning processes.

VIII. ACTIVE LEARNING EXPERIMENTS

Experiments have been carried out on the databases described in the appendix. In the following, we use a SVM classifier with a Gaussian kernel and a χ^2 distance.

A. Comparison of active learners

We compare the method proposed in this paper to an uncertainty-based method SVM_{active} [33], and a method which aims at minimizing the error of generalization [26]. We also add a non active method, which randomly selects the images.

Results per concept are shown in Table I for the Corel photo database for 10 to 50 labels. First of all, one can see that we have selected concepts of different levels of complexities. The performances go from few percentages of Mean Average
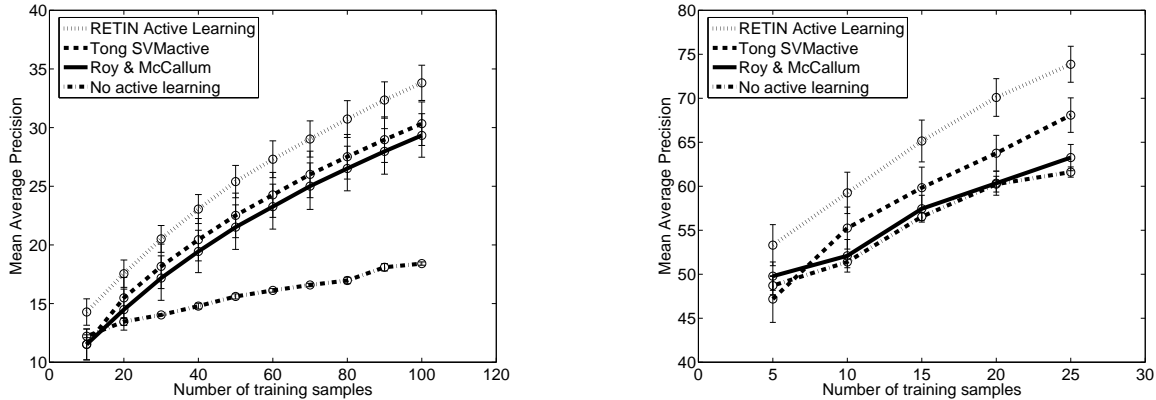
Fig. 4. Mean Average Precision(%) for different active learners on the Corel (left) and the ANN (right) photo database.

Precision to 89%. The concepts that are the most difficult to retrieve are very small and/or have a very diversified visual content. For instance, the "Ontario" concept has 42 images, and is compound of pictures of buildings and landscapes with no common visual content. However, the system is able to make some improvements. Next, the RETIN active learning method has the best performances for most concepts.

Global results are shown in Fig. VII. First, one can see the benefit of active learning in our context, which increases MAP from 11% to 15% for the Corel database. The method which aims at minimizing the error of generalization is the less efficient active learning method. The most efficient method is the precision-oriented method we introduced in this paper, especially in the first iterations, where the number of samples is smallest. About computational time per feedback, the SVM_{active} method needs about 20ms, the method of [26] several minutes, and the one proposed in this paper 45ms on the Corel photo database.

One can also notice that the gap between active and non-active learning is larger on the Corel photo database, which has 10 times more images than the ANN database. This shows that active learning becomes more and more efficient when the size of the database grows.

B. RETIN Components

The RETIN selection process is composed of three components : Active Boundary Correction (section V), Precision-Oriented Selection (section VI), and Diversification (section VII). Experiments have been carried out on the Corel database in order to show the level of improvement brought by each of these components. The results are shown in Fig. VIII-C. The top curve shows the performances using all the components, the next one using Diversification and Active Boundary Correction, the next one only using Active Boundary Correction, and the last one is the selection of the images the closest to the boundary, with no components.

The first improvements come from the Active Boundary Correction and Diversification, but in different ways. The boundary correction increases the performances the most in the first feedback steps, while the diversification increases the performances the most after several feedback steps. This behavior is as expected, since the boundary correction aims at handling the lack of training samples during the first feedback steps, and the diversification is more interesting when the category becomes large. Finally, the top curve shows that the enhancement of the SVM_{active} criterion proposed in section VI increases the performances when used with correction and diversification.

C. Labeling system parametrization

We ran simulations with the same protocol that in the previous section, but with various numbers of labels per feedback. In order to get comparable results, we ensure that the size of the training set at the end of a retrieval session is always the same. We compute the precision/recall curves for all the concepts of the database. Results for the "savanna" concept are shown in Fig. VIII-C; let us note that all concepts gave similar results moduling a scaling factor. As one can see on this figure, the more feedback steps, the more performances increase. Increasing feedback steps leads to more classification updates, which allows a better correction and selection.

IX. CONCLUSION

In this paper, the RETIN active learning strategy for interactive learning in Content-Based Image Retrieval context is presented. The classification framework for CBIR is studied and powerful classification techniques for information retrieval context are selected. After analyzing the limitation of active learning strategies to the CBIR context, we introduce the general RETIN active learning scheme, and the different components to deal with this particular context. The main contributions concern the boundary correction to make the retrieval process more robust, and secondly, the introduction of a new criterion for image selection that better represents the CBIR objective of database ranking. Other improvements, as batch processing and speed-up process are proposed and discussed. Our strategy leads to a fast and efficient active learning scheme to online retrieve query concepts from a database. Experiments on large databases show that the RETIN
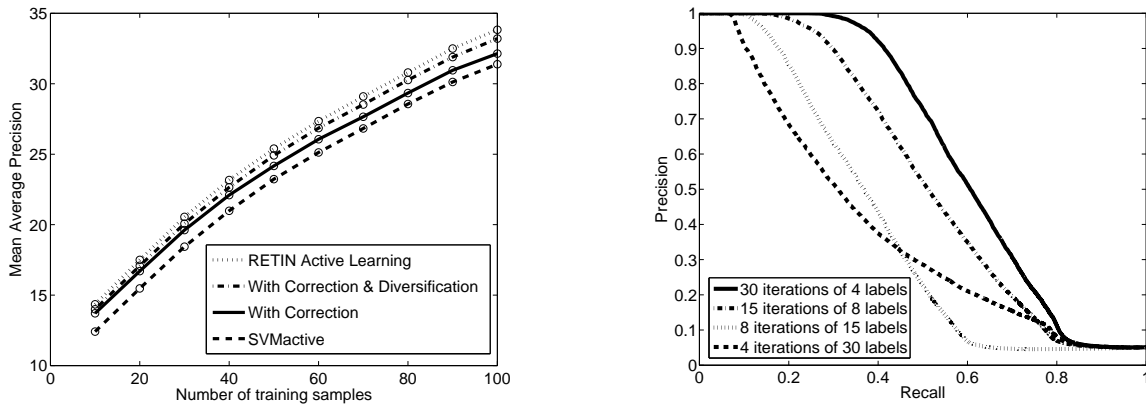
Fig. 5. Results on the Corel photo database. (left) Mean Average Precision(%) using one to three components of the RETIN selection process. (right) Precision/Recall curves for the concept 'savanna'.

method gives very good results in comparison to several other active strategies.

The framework introduced in this article may be extended. We are currently working on kernel functions for object classes retrieval, based on bags of features: each image is no more represented by a single global vector, but by a set of vectors. The implementation of such a kernel function is fully compatible with the RETIN active learning scheme described in this article, and the initial results are really encouraging.

APPENDIX: DATABASES AND FEATURES FOR EXPERIMENTS

Databases. We considered two databases. The first one is the COREL photo database. To get tractable computation for the statistical evaluation, we randomly selected 77 of the COREL folders, to obtain a database of $6,000$ images. The second one is the ANN database from the Washington university, with 500 images.

Features. We use a histogram of 25 colors and 25 textures for each image, computed from a vector quantization [35].

Concepts. For the ANN database, we used the already existing 11 concepts. For the Corel database, in order to perform interesting evaluation, we built from the database 50 concepts of various complexity. Each concept is built from 2 or 3 of the COREL folders. The concept sizes are from 50 to 300. The set of all concepts covers the whole database, and many of them share common images.

Quality assessment. The CBIR system performance measurement is based on the Precision and Recall.

Precision and Recall are considered for one category, and have as many values as images in the database (from one image retrieved to the maximum number of images the system can return). Precision and Recall are metrics to evaluate the ranking of the images returned by the system for one category. The Precision curve is always decreasing (or stationary), and the best Precision curve is the one which decreases the less, which means that whatever the number of images retrieved by the system, a lot of them are relevant ones. The Recall curve is always increasing, and the best Recall curve is the one which increases the fastest, which means that the system

has retrieved most of the relevant images, and few of them are lost.

Precision and Recall are interesting for a final evaluation of one category, however for larger evaluation purposes, we consider the Precision/Recall curve. This curve is the set of all the couples (Precision, Recall) for each number of images returned by the system. The curve always starts from the top left (1,0) and ends in the bottom right (0,1). Between these two points, the curve decreases regularly. A good Precision/Recall curve is a curve which decreases slowly, which means that at the same time, the system returns a lot of relevant images and few of them are lost. This property is interesting since the size of the category is not playing an important role, which allows the comparison of Precision/Recall for different categories.

The Precision/Recall curve can also be summarized by a single real value called Average Precision, which corresponds to the area under an ideal (non-interpolated) recall/precision curve.

To evaluate a system over all the categories, the Average Precisions for each category are combined (averaged) across all categories to create the non-interpolated Mean Average Precision (MAP) for that set. Let's note that this criterion is the one used by the TRECVID evaluation campaign[37].

The Precision and Recall values are measured by simulating retrieval scenario. For each simulation, an image category is randomly chosen. Next, 100 images are selected using active learning and labeled according to the chosen category. These labeled images are used to train a classifier, which returns a ranking of the database. The Average Precision is then computed using the ranking. These simulations are repeated 1000 times, and all values of are averaged to get the Mean Average Precision. Next, we repeat ten times these simulations to get the mean and the standard deviation of the MAP.

RETIN interface.

The RETIN user interface (*cf.* Fig. 6) is composed of three sub-parts. The main one at the top left displays the current ranking of the database. The second at the bottom displays the current selection of the active learner. The user gives new labels by clicking the left or right mouse button. Once new labels are given, the retrieval is updated, and a new
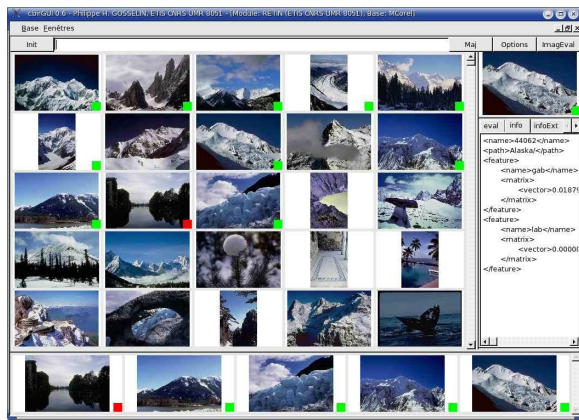
Fig. 6.   RETIN user interface.

ranking is displayed in the main part. The last part at the top right displays information about the current image. We show on Fig. 7 the 75 most relevant pictures after 3 and 5 iterations of 5 labels for the concept "mountain". One can see that the system is able to retrieve many images of the concept, while discriminating irrelevant pictures with close visual characteristics as for example pictures of dolphins. This concept is very complex and overlaps with several other concepts and will necessitate more iterations to be completely learned. Online demonstration is also available on our website (http://retin.ensea.fr).

REFERENCES

[1] R. Veltkamp, "Content-based image retrieval system: A survey," University of Utrecht, Tech. Rep., 2002.

[2] Y. Rui, T. Huang, S. Mehrotra, and M. Ortega, "A relevance feedback architecture for content-based multimedia information retrieval systems," in *IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1997, pp. 92–89.

[3] E. Chang, B. T. Li, G. Wu, and K. Goh, "Statistical learning for effective visual information retrieval," in *IEEE International Conference on Image Processing*, Barcelona, Spain, September 2003, pp. 609–612.

[4] S. Aksoy, R. Haralick, F. Cheikh, and M. Gabbouj, "A weighted distance approach to relevance feedback," in *IAPR International Conference on Pattern Recognition*, vol. IV, Barcelona, Spain, September, 3-8 2000, pp. 812–815.

[5] J. Peng, B. Bhanu, and S. Qing, "Probabilistic feature relevance learning for content-based image retrieval," *Computer Vision and Image Understanding*, vol. 75, no. 1-2, pp. 150–164, July-August 1999.

[6] N. Doulamis and A. Doulamis, "A recursive optimal relevance feedback scheme for cbir," in *International Conference in Image Processing (ICIP'01)*, Thessaloniki, Greece, October 2001.

[7] J. Fournier, M. Cord, and S. Philipp-Foliguet, "Back-propagation algorithm for relevance feedback in image retrieval," in *International Conference in Image Processing (ICIP'01)*, vol. 1, Thessaloniki, Greece, October 2001, pp. 686–689.

[8] O. Chapelle, P. Haffner, and V. Vapnik, "Svms for histogram based image classification," *IEEE Transactions on Neural Networks*, vol. 10, pp. 1055–1064, 1999.

[9] N. Vasconcelos, "Bayesian models for visual information retrieval," Ph.D. dissertation, Massachusetts Institute of Technology, 2000.

[10] S.-A. Berrani, L. Amsaleg, and P. Gros, "Recherche approximative de plus proches voisins : application la reconnaissance d'images par descripteurs locaux," *Technique et Science Informatiques*, pp. 22(9):1201–1230, 2003.

[11] N. Najjar, J. Cocquerez, and C. Ambroise, "Feature selection for semi supervised learning applied to image retrieval," in *IEEE ICIP*, vol. 2, Barcelona, Spain, Sept. 2003, pp. 559–562.

[12] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *International Conference on Machine Learning*, 2003.

[13] A. Dong and B. Bhanu, "Active concept learning in image databases," *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, vol. 35, pp. 450–466, 2005.

[14] N. Vasconcelos and M. Kunt, "Content-based retrieval from image databases: current solutions and future directions," in *International Conference on Image Processing*, vol. 3, Thessaloniki, Greece, October 2001, pp. 6–9.

[15] Y. Chen, X. Zhou, and T. Huang, "One-class svm for learning in image retrieval," in *International Conference in Image Processing (ICIP)*, vol. 1, Thessaloniki, Greece, October 2001, pp. 34–37.

[16] S.-A. Berrani, L. Amsaleg, and P. Gros, "Approximate searches: k-neighbors + precision," in *International Conference on Information and Knowledge Management*, 2003, pp. 24–31.

[17] V. Vapnik, *Statistical Learning Theory*.   Wiley-Interscience, New York, 1998.

[18] S. Tong and D. Koller, "Support vector machine active learning with application to text classification," *Journal of Machine Learning Research*, pp. 2:45–66, November 2001.

[19] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller, "Fisher discriminant analysis with kernels," in *Neural Networks for Signal Processing IX*, Y.-H. Hu, J. Larsen, E. Wilson, and S. Douglas, Eds.   IEEE, 1999, pp. 41–48. [Online]. Available: citeseer.ist.psu.edu/mika99fisher.html

[20] B. Schölkopf and A. Smola, *Learning with Kernels*.   MIT Press, Cambridge, MA, 2002.

[21] P. Yin, B. Bhanu, K. Chang, and A. Dong, "Integrating relevance feedback techniques for image retrieval using reinforcement learning," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1536–1551, October 2005.

[22] J. Shawe-Taylor and N. Cristianini, *Kernel methods for Pattern Analysis*. Cambridge University Press, ISBN 0-521-81397-2, 2004.

[23] T. Joachims, "Transductive inference for text classification using support vector machines," in *Proc. 16th International Conference on Machine Learning*.   Morgan Kaufmann, San Francisco, CA, 1999, pp. 200–209.

[24] P. Gosselin, M. Najjar, M. Cord, and C. Ambroise, "Discriminative classification *vs* modeling methods in CBIR," in *IEEE Advanced Concepts for Intelligent Vision Systems (ACIVS)*, Brussel, Belgium, September 2004.

[25] P. Gosselin and M. Cord, "RETIN AL: An active learning strategy for image category retrieval," in *IEEE International Conference on Image Processing*, vol. 4, Singapore, October 2004, pp. 2219–2222.

[26] N. Roy and A. McCallum, "Toward optimal active learning through sampling estimation of error reduction," in *International Conference on Machine Learning*, 2001.

[27] D. Cohn, "Active learning with statistical models," *Journal of Artificial Intelligence Research*, vol. 4, pp. 129–145, 1996.

[28] D. D. Lewis and J. Catlett, "Heterogeneous uncertainly sampling for supervised learning," in *International Conference on Machine Learning*. W.W. Cohen and H. Hirsh, editors, July 1994, pp. 148–56.

[29] J. M. Park, "Online learning by active sampling using orthogonal decision support vectors," in *IEEE Workshop on Neural Networks for Signal Processing*, vol. 77, no. 3, December 2000, pp. 263–283.

[30] M. Lindenbaum, S. Markovitch, and D. Rusakov, "Selective sampling for nearest neighbor classifiers," *Machine Learning*, pp. 54(2):125–152, February 2004.

[31] I. Cox, M. Miller, T. Minka, T. Papathomas, and P. Yianilos, "The bayesian image retrieval system, PicHunter: Theory, implementation and psychophysical experiments," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 20–37, 2000.

[32] D. Geman and R. Moquet, "A stochastic feedback model for image retrieval," in *RFIA'2000*, vol. III, Paris, France, February 2000, pp. 173–180.

[33] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *ACM Multimedia*, 2001, pp. 107–118.

[34] P. Gosselin and M. Cord, "A comparison of active classification methods for content-based image retrieval," in *International Workshop on Computer Vision meets Databases (CVDB), ACM Sigmod*, Paris, France, June 2004.

[35] M. Cord, P. Gosselin, and S. Philipp-Foliguet, "Stochastic exploration and active learning for image retrieval," *Image and Vision Computing*, vol. 25, pp. 14–23, 2006.

[36] K. Brinker, "Incorporating diversity in active learning with support vector machines," in *International Conference on Machine Learning*, February 2003, pp. 59–66.

[37] "TREC Video Retrieval Evaluation Campain," http://www-nlpir.nist.gov/projects/trecvid/.
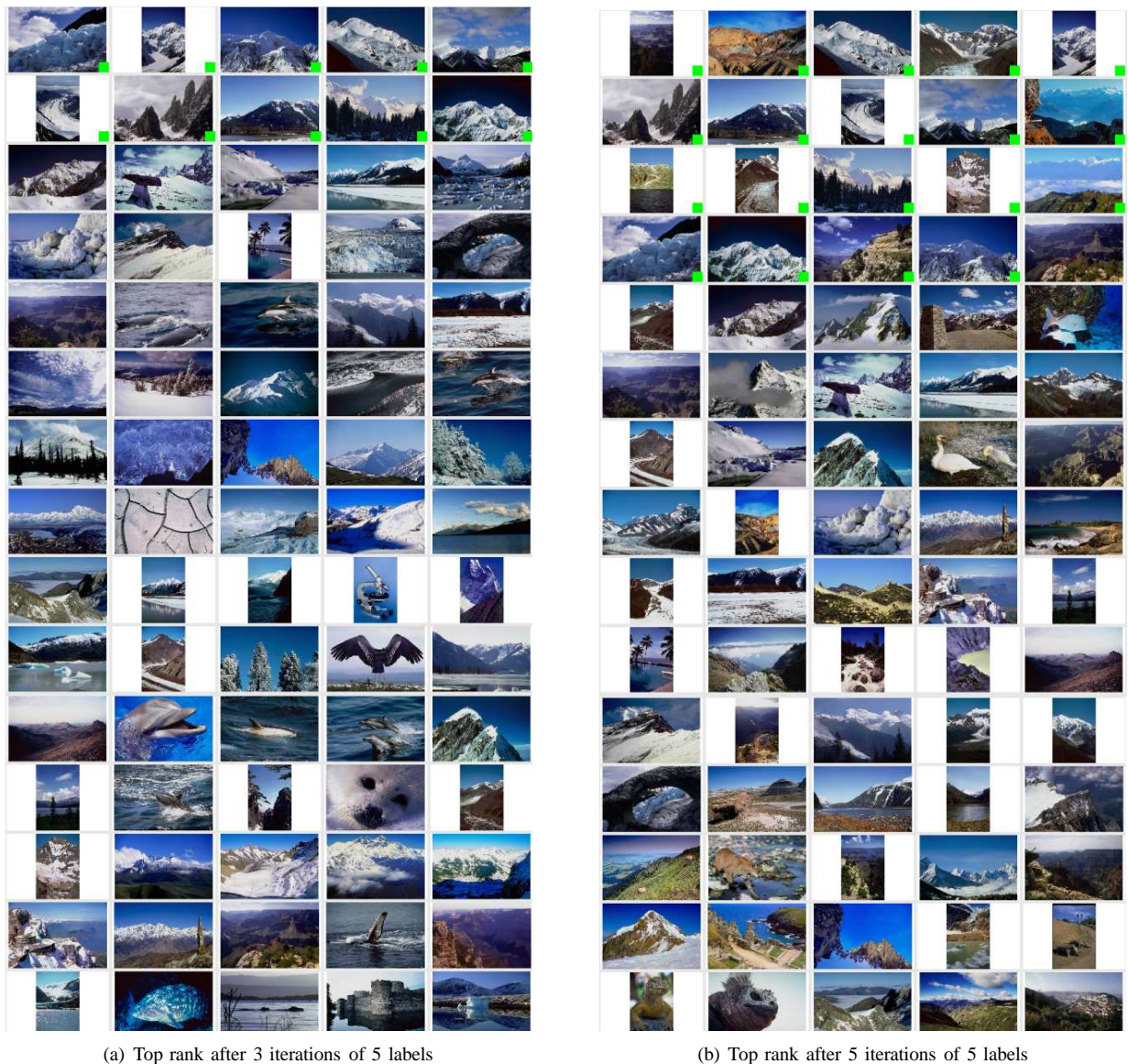
(a) Top rank after 3 iterations of 5 labels

(b) Top rank after 5 iterations of 5 labels

Fig. 7. RETIN results: 75 most relevant pictures for the concept "mountain" after 3 or 5 feedback iterations of our active learning-based strategy.

Philippe Henri Gosselin received the PhD degree in image and signal processing in 2005 (Cergy, France). After 2 years of post-doctoral positions at the LIP6 Lab. (Paris, France) and at the ETIS Lab. (Cergy, France), he joined the MIDI Team in the ETIS Lab as an assistant professor. His research focuses on machine learning for online muldimedia retrieval. He developped several statistical tools to deal with the special characteristics of content-based multimedia retrieval. This include studies on kernel functions on histograms, bags and graphs of features, but also weakly supervised semantic learning methods. He is involved in several international research projects, with applications to image, video and 3D object databases. He is a member of the European network of excellence MUSCLE (Multimedia Understanding through Semantics, Computation and LEarning).

Matthieu Cord received the PhD degree in image processing in 1998 (Cergy Univ., France). After 1 year post-doc position at the ESAT lab. (KUL, Belgium), he joined the ETIS lab., France. He received the HDR degree and got a full professor position at the LIP6 lab. (UPMC, Paris) in 2006. His research interests include Content-Based Image Retrieval and Computer Vision. In CBIR, he focuses on learning-based approaches to visual information retrieval. He developed several interactive systems, to implement, compare and evaluate online relevance feedback strategies for web-based applications and artwork dedicated contexts. He is also working on multimedia content analysis. He is involved in several international research projects. He is a member of the European network of excellence MUSCLE (Multimedia Understanding through Semantics, Computation and LEarning).

TABLE I
MEAN AVERAGE PRECISION(%) PER CONCEPT FOR EACH ACTIVE LEARNER, ON THE COREL PHOTO DATABASE. (A) = ROY & MCCALLUM MINIMIZATION OF THE ERROR OF GENERALIZATION; (B) = SVM$_{active}$; (C) = RETIN ACTIVE LEARNING.

Method	10 labels			20 labels			30 labels			40 labels			50 labels		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
animals	29	39	**41**	34	41	**45**	41	43	**49**	46	45	**52**	49	45	**54**
caves	20	21	**22**	24	26	**32**	28	28	**44**	32	32	**56**	39	37	**61**
doors	27	33	**41**	40	47	**54**	51	53	**63**	56	54	**67**	60	56	**70**
fish	50	55	**66**	62	66	**79**	69	72	**85**	73	75	**88**	78	77	**89**
flowers	18	23	**37**	25	27	**43**	31	29	**46**	35	32	**48**	38	33	**49**
food	38	37	**47**	53	54	**58**	59	59	**65**	62	61	**69**	64	65	**72**
rome	17	16	**20**	24	24	**30**	33	32	**36**	38	37	**43**	41	41	**48**
roses	27	31	**38**	36	37	**47**	45	41	**53**	49	44	**59**	54	48	**63**
african	**8**	**8**	**8**	9	9	**11**	10	12	**13**	12	14	**15**	13	15	**17**
savana	14	17	**18**	19	**23**	22	23	**29**	27	27	**32**	31	31	**35**	35
asia	5	5	**6**	7	7	**8**	9	**10**	**10**	10	**12**	**12**	13	14	**15**
bears	7	8	**10**	9	10	**13**	12	13	**15**	13	15	**17**	14	17	**18**
birds	6	6	**7**	6	7	**9**	7	8	**10**	8	9	**12**	9	10	**14**
city	10	9	**11**	**14**	13	14	18	17	**19**	21	20	**22**	23	23	**24**
cougars	**8**	6	**8**	**12**	8	10	**14**	10	12	**17**	12	13	**19**	15	16
country	9	8	**10**	9	9	**10**	10	10	**11**	11	11	**12**	12	12	**13**
desert	7	**8**	**8**	8	**9**	**9**	9	**10**	**10**	10	**11**	**11**	11	**12**	**12**
dogs	8	9	**15**	12	12	**19**	16	14	**24**	21	17	**28**	24	19	**32**
england	7	6	**8**	**8**	7	**8**	**9**	8	**9**	**10**	9	**10**	**11**	10	**11**
europe	7	8	**9**	8	9	**10**	8	9	**10**	9	10	**11**	10	**11**	**11**
fabulous	8	**9**	8	11	**12**	**12**	14	**15**	**15**	16	17	**18**	19	19	**21**
festive	19	21	**24**	26	27	**32**	31	30	**37**	35	33	**41**	38	34	**43**
finland	8	8	**9**	9	9	**11**	11	12	**13**	14	**16**	15	16	17	**18**
france	17	22	**24**	21	26	**27**	25	29	**30**	27	**31**	31	29	33	**34**
fruit	11	11	**12**	15	16	**19**	19	21	**25**	24	25	**32**	27	30	**37**
garden	9	8	**12**	11	11	**15**	14	15	**18**	16	18	**21**	18	20	**23**
greece	4	5	**6**	5	6	**8**	6	7	**10**	7	9	**12**	9	10	**14**
house	**6**	**6**	**6**	7	7	**8**	**10**	9	**10**	11	10	**12**	12	12	**13**
ice	8	**12**	10	9	**12**	**12**	11	13	**14**	12	15	**16**	14	17	**18**
meditar.	8	**10**	9	10	**13**	11	13	**17**	14	17	**21**	17	19	**24**	20
mountains	10	11	**15**	13	15	**18**	17	18	**21**	20	22	**23**	24	**26**	**26**
objects	11	11	**12**	14	14	**19**	20	20	**25**	24	24	**29**	29	28	**33**
people	5	6	**7**	6	**8**	**8**	8	8	**9**	8	9	**10**	9	10	**11**
savana	23	22	**26**	29	29	**31**	**34**	**34**	**34**	**38**	37	**38**	**42**	39	**42**
sunsets	11	14	**16**	16	22	**24**	21	27	**31**	25	32	**36**	28	36	**41**
tropical	5	6	**7**	7	8	**9**	9	10	**11**	11	12	**14**	14	13	**16**
usa	8	**12**	10	9	**12**	11	10	**12**	**12**	11	12	**13**	11	12	**14**
wine	4	5	**6**	6	7	**8**	7	**9**	**9**	8	10	**11**	10	12	**13**
beaches	4	**5**	**5**	5	5	**6**	5	**7**	6	6	**8**	**8**	7	**8**	**8**
britain	**4**	**4**	**4**	4	**5**	4	**5**	**5**	**5**	5	**6**	**6**	**6**	**6**	**6**
canada	**3**	**3**	**3**	**3**	**3**	**3**	3	3	**4**	3	**4**	**4**	4	4	**5**
castles	**6**	**6**	5	**6**	**6**	**6**	**7**	**7**	**7**	7	**8**	**8**	8	**9**	**9**
green	3	3	**4**	3	4	**5**	4	5	**6**	5	5	**7**	5	7	**8**
india	4	4	**5**	5	5	**7**	6	5	**8**	7	6	**9**	8	7	**10**
kenya	4	**5**	**5**	5	**6**	**6**	6	**8**	7	8	**10**	8	9	**11**	10
ontario	3	3	**4**	3	3	**4**	**4**	**4**	**4**	4	4	**5**	**5**	**5**	**5**
portugal	3	3	**4**	3	3	**4**	4	4	**5**	4	**5**	**5**	4	5	**6**
underwater	3	**4**	**4**	4	**5**	**5**	5	**6**	**6**	6	**7**	6	6	**8**	7
europe	**5**	**5**	**5**	5	5	**6**	**6**	**6**	**6**	6	6	**7**	7	7	**8**
Mean	11	12	**14**	14	15	**18**	17	18	**21**	19	20	**23**	21	22	**25**