

# PRECISION-ORIENTED ACTIVE SELECTION FOR INTERACTIVE IMAGE RETRIEVAL

Philippe H. Gosselin and Matthieu Cord

ETIS / CNRS UMR 8051  
6, avenue du Ponceau, 95014 Cergy-Pontoise, France

## ABSTRACT

Active learning methods have been considered with an increased interest in the content-based image retrieval (CBIR) community. These methods have been developed for classification problems, and do not deal with the particular characteristics of the CBIR. One of these characteristics is the criterion to optimize, for instance the error of generalization for classification, which is not the best adapted to CBIR context. We introduce in this paper an active selection which chooses the image the user should label such as the Mean Average Precision is increased. The method is smartly combined with previous propositions, and leads to a fast and efficient active learning scheme. Experiments on a large database have been carried out in order to compare our approach to several other methods.

**Index Terms**— Image classification, Image databases, Learning systems, Information retrieval

## 1. 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, focused on fully automatic strategies, recent approaches introduce human-computer interaction [1, 2].

Starting with a coarse query, the interactive process allows the user to refine his request as much as necessary. Many kinds of interaction between the user and the system have been proposed [3], but most of the time, user information consists of binary annotations (labels) indicating whether or not the image belongs to the desired category.

In this paper, we focus on the retrieval of *concepts* within a large document collection. We assume that a user is looking for a set of documents, the query concept, within an existing document database. The aim is to build a fast and efficient strategy to retrieve the query concept.

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 [4]. The relevant class is the set of documents within the query concept, and the irrelevant class the set of documents out of the query concept.

Active learning has been introduced with success in CBIR context in order to deal with the interaction between the user and the system [5]. However, a lot of active strategies consider CBIR as a pure classification problem, and thus are not fully adapted to the special characteristics of this context. For instance, we have shown in a previous paper [6] that the few training data and the imbalance of the classes lead to a noisy boundary, although its quality is an important factor in active classification.

In this paper, we focus on another characteristic of CBIR : the active selection criterion. Active classification aims at minimizing the error of classification. However, this criterion is not the most

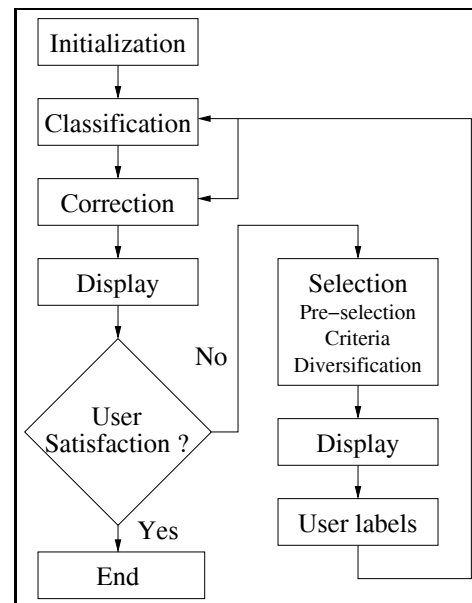


Fig. 1. Active learning scheme for interactive image retrieval

representative of user satisfaction. Thus, we propose to select images the user should label using common metric of CBIR, the Precision.

In section 2, we present a global active learning scheme for interactive CBIR. In section 3, we discuss about the differences between active selection based on error of classification and Average Precision. In section 4, we introduce our active learning scheme which aims at maximizing the Precision-oriented criterion. In section 5, experiments are carried out on a general image database in order to compare different active learning strategies.

## 2. ACTIVE LEARNING SCHEME

We propose an active learning scheme based on binary classification in order to interact with a user searching for an image concept in the database. That scheme is summarized in figure 1. We denote by  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  the images, and by  $\mathbf{y} = \{y_1, \dots, y_n\}$  the user labels (1 if relevant, -1 if irrelevant, 0 if unlabeled). The examples are the images  $i \in I$  with a non-zero label, *i.e.* couples  $(\mathbf{x}_i, y_i)$  where  $y_i \neq 0$ ,

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

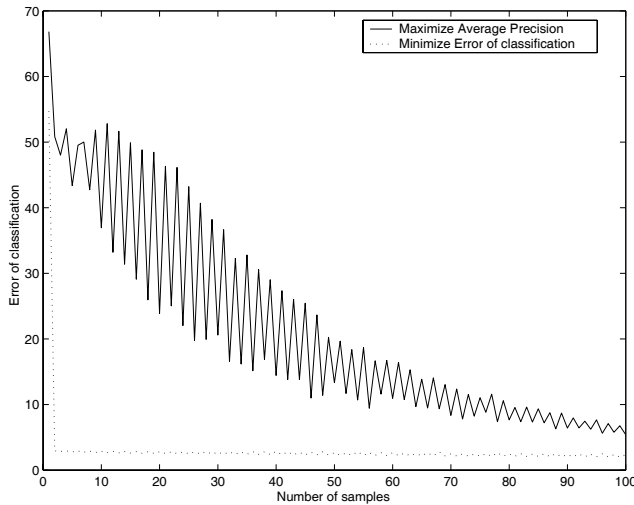


Fig. 2. Error of classification with two “perfect” active learners.

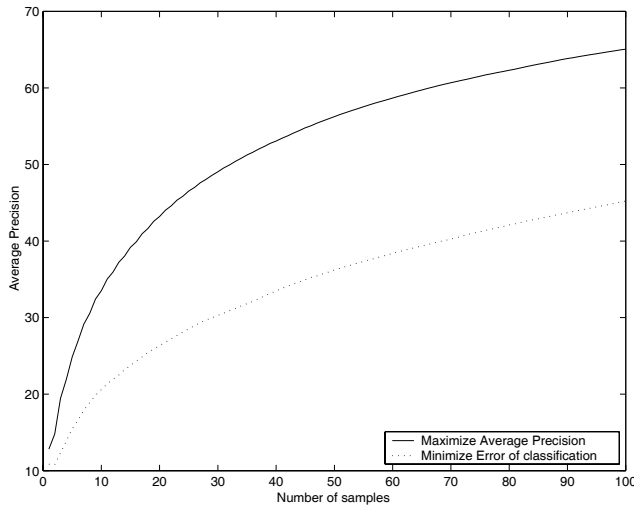


Fig. 3. Mean Average Precision with two “perfect” active learners.

**Classification.** A binary classifier is trained with the examples the user has given. In this paper, we use a SVM with a gaussian  $\chi^2$  kernel, as it has been revealed as the most efficient [5, 7]. The result is a function  $f_{\mathbf{y}}(\mathbf{x})$  which returns the relevance of each image  $\mathbf{x}$ , after a training with the examples  $(\mathbf{x}_i, y_i), i \in I$ .

**Correction.** We add an active correction to the class boundary in order to deal with the small number of training samples and the imbalance of the classes (the size of the searched category is usually very small compared to the number of irrelevant images). Details about this method can be found in [6].

**Selection.** The main purpose of this paper comes in this part. When the user is not satisfied with the current classification, the system selects a batch of images the user should label. The selection will be such as the labeling of those images will give the best performances. We divide the selection in three steps.

The first step aims at reducing the computational time, by pre-selecting some hundreds of pictures which may be in the optimal selection set. We propose to pre-select a set indexed by  $J$  of the closest

pictures to the (corrected) boundary. This process is computed very fast, and the uncertainly-based selection method has proven its interest in CBIR context.

The second step is the computation of the selection criterion. In active classification, the criterion is the minimization of the error of classification (or *risk*). In these cases, the risk is computed for each classification function  $f_{\mathbf{y}, t(\mathbf{x}_i)}$ , which is trained with the label  $t(\mathbf{x}_i)$  of an unlabeled image  $i \notin I$  added to current training set  $(\mathbf{y})$ . Finally, the selected image  $i^*$  is the one which minimizes the risk:

$$i^* = \operatorname{argmin}_{i \notin I} \operatorname{risk}(f_{\mathbf{y}, t(\mathbf{x}_i)})$$

The main difficulty of this task is the fact that the label  $t(\mathbf{x}_i)$  is unknown, and an estimation is required. This estimation is replaced with a *cost* function that we denote by  $g_{\mathbf{y}}(\mathbf{x}_i)$ , and including the pre-selection, the problem can be written as:

$$i^* = \operatorname{argmin}_{i \in J} g_{\mathbf{y}}(\mathbf{x}_i)$$

The cost function is, for instance, the distance to the boundary for the SVM<sub>active</sub> method [5]:  $g_{\mathbf{y}}(\mathbf{x}_i) = |f_{\mathbf{y}}(\mathbf{x}_i)|$ . In this paper, we introduce a cost function  $g_{\mathbf{y}}(\mathbf{x}_i)$  in this scheme, aiming at maximizing the Average Precision.

The third step of active selection computes the batch selection. As we focus on real-time application, we use a fast method close to the angle diversity [8]. The method selects  $q$  images using the previously computed cost  $g_{\mathbf{y}}(\mathbf{x}_i)$ , and return the set  $I^*$  of image indexes the user should label:

$$\begin{aligned} I^* &= \{\} \\ \text{for } l &\in [1..q] \\ i^* &= \operatorname{argmin}_{i \in J - I^*} (g_{\mathbf{y}}(\mathbf{x}_i) + \max_{j \in I \cup I^*} s(\mathbf{x}_i, \mathbf{x}_j)) \\ I^* &= I^* \cup \{i^*\} \\ \text{endfor} \end{aligned}$$

where  $s(\mathbf{x}_i, \mathbf{x}_j)$  is the similarity (using the kernel) between images  $\mathbf{x}_i$  and  $\mathbf{x}_j$ .

**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.

### 3. ERROR OF GENERALIZATION VS MEAN AVERAGE PRECISION (MAP)

Active classification methods have been built to select elements which decrease the error of classification. However, in our interactive CBIR context, users are concerned by a similarity ranking of the database. An usual metric to evaluate this ranking is the Mean Average Precision<sup>1</sup>, *i.e.* the sum of the Precision/Recall curve.

Active classification methods have already proven their capacity to increase the Average Precision, whenever this is not what they aim at. In order to evaluate the two criteria, – classification error-based minimization and MAP-based maximization, we propose an experimentation using a ground-truth.

Active learning methods aim at optimizing some criterion, but in practice the true evaluation of the criterion is impossible, because we do not have the ground truth. Thus, active classification methods differ one to another by the way they estimate the error of classification. In this section, we propose to use two “perfect” active learner,

<sup>1</sup>cf. TREC VIDEO conference: <http://www-nlpir.nist.gov/projects/trecvid/>

with no estimation, in order to compare the two criteria. These methods are not designed for real application, and use the ground truth to compute the error of classification or the Average Precision.

The results of those experiments are shown on Figure 2 and 3. Considering error of classification (*cf.* Fig. 2), the method which directly optimizes the error criterion is the most efficient. However one can see on this figure that maximizing the Average Precision also decreases the error. Considering the criteria that interest us (*cf.* Fig. 3), we can see a similar but inverted behavior. The most interesting result is the fact that, although the minimization of the error of classification increases the Average Precision, direct maximization of the Average Precision leads to a large increasing of performance, with a gain of 20%.

#### 4. PRECISION-ORIENTED SELECTION

The results of the previous experiments lead us to the exploration of a precision-oriented selection. As an active classification method has to minimize the error of classification, the aim here is to select the images so that the Average Precision will be increased. Furthermore, we also need to perform an estimation of the criterion.

However, estimating the Average Precision is particularly difficult with the kind of samples we have chosen. As we work on generalist CBIR systems, we opted for binary labels, which are simple enough for any non-expert user. These labels are well adapted for classification, but are less suitable with the estimation of Average Precision : the criterion is based on a ranking of the database, but binary labels do not give any information in that way.

Researchers propose to use true/false positives/negatives in order to solve this problem. However, in our context, the number of samples is so small that it becomes especially difficult to find such an information. In other words, the labels are so precious that each one will necessarily have a huge weight in the result.

Thus, we propose to stay in an active classification framework, and introduce a bias in the selection in order to support the increasing of the Average Precision. We weight the distance to the boundary by a factor  $h_{\mathbf{y}}(\mathbf{x}_i)$ :

$$g_{\mathbf{y}}(\mathbf{x}_i) = |f_{\mathbf{y}}(\mathbf{x}_i)| \times (1 - h_{\mathbf{y}}(\mathbf{x}_i))$$

The aim of introducing  $h_{\mathbf{y}}(\mathbf{x}_i)$  is to support the picture  $\mathbf{x}_i$  so that, once labeled, the Average Precision is increased. In order to get this behavior, we propose to consider the subset of the labeled pictures. As we have the labels of all images of the subset, we can use it as a ground truth. Thus, it becomes feasible to compute the Average Precision on this subset, without any estimation. We still need a ranking of this subset in order to compute the Average Precision. We propose to compute the similarity of an unlabeled image  $\mathbf{x}_i$  to all the labeled images, and then rank the labeled images according to these similarities. The resulting factor  $h_{\mathbf{y}}(\mathbf{x}_i)$  is the Average Precision on the subset with this ranking. Note that we could use the ranking provided by the classifier. However, we have experimented this approach, and the results are not good still because of the few training data and the binary labels.

#### 5. EXPERIMENTS

We experimented different active learning methods with the following protocol :

**Database.** The image database is an extract of 6,000 images of the COREL photo database. To get tractable computation for the statistical evaluation, we randomly selected 77 of the COREL folders.

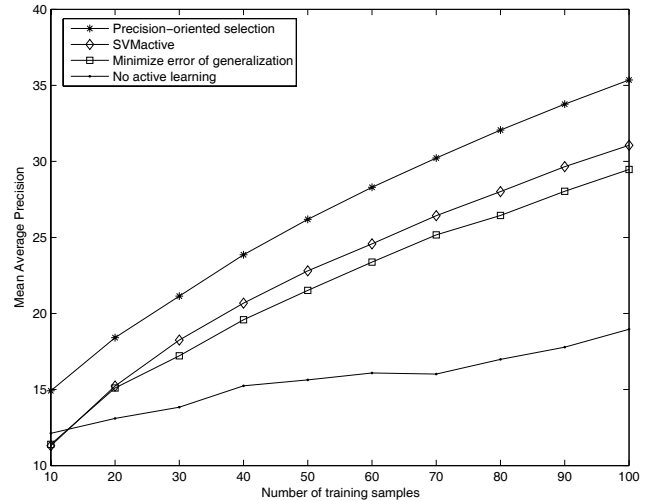


Fig. 4. Mean Average Precision(%) for different active learners.

**Features.** We use an histogram of 25 colors and 25 textures for each image, computed from a vector quantization [9].

**Classifier.** We use SVM classifier with a gaussian kernel with a  $\chi^2$  distance.

**Concepts.** To perform interesting evaluation, we built from the database 50 concepts. Each concept is made of 2 or 3 of the COREL folders. The concept sizes are from 50 to 300. The set of all the concepts covers the whole database, and many of them share common images.

**Active methods.** We compare the method proposed in this paper to an uncertainty base method :  $SVM_{active}$  [5], and a method which aims at minimizing the error of generalization [10]. We also add a non active method, which randomly selects the images.

The results of the experiments are shown in Fig. 4. Firstly, one can see the benefit of active learning in our context. In these experiments, the gain is from 11% to 17%. 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 introduce in this paper, especially in the first iterations, where the number of samples is the smallest. About computational time per feedback, the  $SVM_{active}$  method need at most 22ms, the method of [10] several minutes, and the one proposed in this paper at most 45ms.

All these methods are exploited by real users through the RETIN 2 User Interface (*cf.* Fig. 5). The interface is compound of three sub-parts. The main one at the top left displays the current ranking of the database. For instance on Fig. 5, we can see the closest pictures to the one brought by the user (top left, with a small green square). The second at the bottom displays the current selection of the active learner. The user can give new labels by clicking the left or right mouse button. Once new labels are given, the user can ask for an update, and the new ranking is displayed in the main part. The last part at the top right displays several informations about an image.

We show on Fig. 6 the 50 most relevant pictures after 5 iterations of 5 labels for the concept “roses”, starting with the query of Fig. 5. One can see that the system is able to retrieve the images of the concept, while discriminating pictures with close visual characteristics. For instance, several non-rose pictures with very close

colors and textures at the beginning of the search (*cf.* Fig. 5) are no more high-ranked 5 iterations later, while the relevant ones are still present (*cf.* Fig. 6).

## 6. CONCLUSION

In this paper, we showed the interest of active learning based on an Average Precision maximization criterion for interactive image retrieval. We introduced a selection method in that way, and combine it with other techniques in a global active learning scheme. The method has been validated through experiments and compared to reference active learning methods. The results show that it is a powerful tool to improve the performances of image category retrieval tasks.

## 7. REFERENCES

- [1] R.C. Veltkamp, "Content-based image retrieval system: A survey," Tech. Rep., University of Utrecht, 2002.
- [2] N. Vasconcelos and M. Kunt, "Content-based retrieval from image databases: current solutions and future directions," in *International Conference in Image Processing*, Thessaloniki, Greece, October 2001, vol. 3, pp. 6–9.
- [3] E. Chang, B. T. Li, G. Wu, and K.S. Goh, "Statistical learning for effective visual information retrieval," in *IEEE International Conference on Image Processing*, Barcelona, Spain, September 2003.
- [4] O. Chapelle, P. Haffner, and V. Vapnik, "Svms for histogram based image classification," *IEEE Transactions on Neural Networks*, vol. 9, 1999.
- [5] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *ACM Multimedia*, 2001.
- [6] P.H. Gosselin and M. Cord, "RETIN AL: An active learning strategy for image category retrieval," in *IEEE International Conference on Image Processing*, Singapore, October 2004.
- [7] P.H. 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.
- [8] K Brinker, "Incorporating diversity in active learning with support vector machines," in *International Conference on Machine Learning*, February 2003, pp. 59–66.
- [9] S. Philipp-Foliguet, G. Logerot, P. Constant, P.H. Gosselin, and C. Lahanier, "Multimedia indexing and fast retrieval based on a vote system," in *IEEE International Conference on Multimedia and Expo*, 2006.
- [10] N Roy and A McCallum, "Toward optimal active learning through sampling estimation of error reduction," in *International Conference on Machine Learning*, 2001.



**Fig. 5.** RETIN 2 User Interface. Center part: retrieved images; Right part: miscellaneous informations about an image; Bottom part: images selected by active learning.



**Fig. 6.** 50 most relevant pictures after 5 iterations of 5 labels for the concept "roses". A small green square indicates an image labeled as relevant.