RETIN AL: AN ACTIVE LEARNING STRATEGY FOR IMAGE CATEGORY RETRIEVAL

Philippe H. Gosselin *Matthieu Cord*

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

ABSTRACT

Active learning methods have been considered with an increasing interest in the content-based image retrieval (CBIR) community. In this article, we propose an efficient method based on active learning strategy to retrieve large image categories. At each feedback step, the system optimizes the image set presented to the user in order to speed up the retrieval. Experimental tests on COREL photo database have been carried out.

1. INTRODUCTION

Content-Based Image Retrieval has attracted a lot of research interest in recent years. Contrary to the early systems, focused on fully automatic strategies, recent approaches introduce human-computer interaction into CBIR [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 large categories, starting with some relevant images brought by the user. Performing an estimation of the searched category can be seen as a statistical learning problem, and more precisely as a binary classification task between the relevant and irrelevant classes [4]. However, the CBIR context defines a very specific learning problem with the following characteristics:

1- *Few training data*. At the beginning, the system has to perform a good estimation of the searched category with very few data. Furthermore, the system can not ask user to label thousands of images, good performances are required using a small percentage of labelled data.

2- *Active Learning*. Due to user annotations, the training data set grows step by step during the retrieval session, so the current classification depends on the previous ones.

We present a method to exploit to the full extent these specificities in order to speed up the retrieval. As Joachims does for text retrieval [5], we propose to use unlabelled data to improve classification when only few training data are available. In image retrieval, annotations are scarce and precious, thus the system has to elicit the user to make them efficiently. We propose an active learning method to select the most difficult images to classify, and to reduce the redundancy in the training data set.

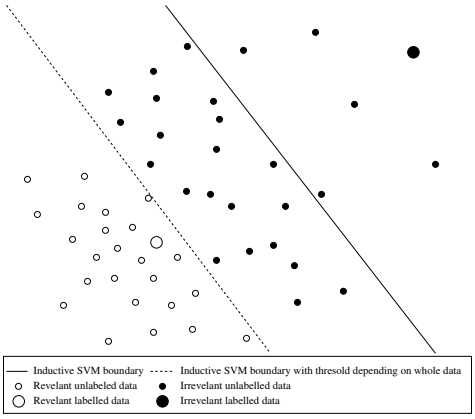

Fig. 1. Few data artifact with SVM boundary.

2. UNLABELLED DATA

As we told in the introduction, the system has to classify the database using only few training data. In the same time, a large amount of unlabelled data is available. If data is structured, unlabelled data may be useful for classification. When a SVM classifier is used, some improvements have been proposed considering unlabelled data. SVM classifier has a decision function f such as:

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y_i k(\mathbf{x}, \mathbf{x}_i) + b \qquad (1)$$

where the $\mathbf{x}_i$ are the feature vectors relative to labelled images, the y_i are the corresponding labels and $k(.,.)$ is a kernel function. α_i and b parameters are computed, considering the SVM optimization.

When very few labels are available, inductive SVM classification may have unexpected results. Fig. 1 shows such a case. Using only labelled data (full line), many irrelevant data are misclassified! Such a configuration may happen when learning samples do not accurately represent the structure of data.

LeSaux[6] proposes to adapt the SVM scheme using unlabelled data. Only one parameter (threshold b) is modified for the whole data. In the case of Fig. 1, this method provides a better classification (dotted line), but in the more complex case of Fig. 2, the boundary does not change (full

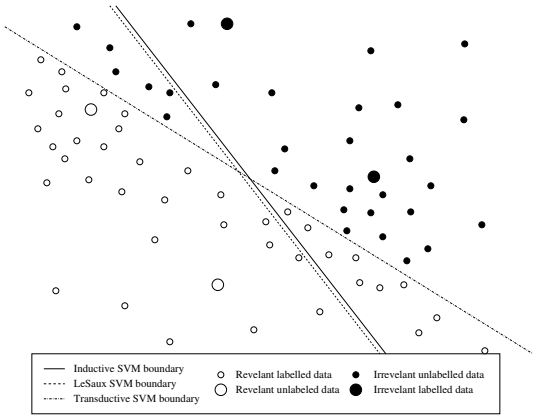

Fig. 2. Transductive SVM.

line). Joachims proposes a method to deal with case of Fig. 2: Transductive SVM [5]. In this particular case, TSVM provides a good classification (dotted line). Proposed in a text retrieval context, we adapt this approach to CBIR.

3. SET SELECTION FOR LABELLING

Performances of inductive classification depend on the training data set. In both previous cases (Fig 1 and 2), other training sets should have provide better classifications. In CBIR, all the images proposed to the user for labelling are added to the training set used for classification. Instead of asking the user to label a random set of images, the active learning tries to focus the user on those whose classification is difficult [7].

3.1. SVM Active Learning

The most known method is SVM$_{active}$ [8]. This method asks user to label twenty images closest to the SVM boundary. Let $(I_j)_{j\in[1..n]}$ be the images of the database, and $r(i,k)$ be the function that, at step i, codes the position k in the relevancy ranking for class appurtenance, using function f of distance to boundary (Eq. 1). At the feedback step i, SVM$_{active}$ proposes to label m images from rank $s(i)$ to $s(i)+m$:

$$\underbrace{I_{r(i,1)}}_{\text{most relevant}}, I_{r(i,2)}, \dots \underbrace{I_{r(i,s(i))}, \dots I_{r(i,s(i)+m-1)}}_{\text{images to label}}, \dots \underbrace{I_{r(i,n)}}_{\text{less relevant}}$$

In SVM$_{active}$ strategy, $s(i)$ is such as $I_{r(i,s(i))}, \dots I_{r(i,s(i)+m-1}$ are the closest images to the SVM boundary. The more an image is close to the margin, the less its classification is accurate. Thus, such images have equal chance of being labelled relevant or irrelevant by the user.

This strategy relies on a strong theoretical foundation and increases performances, but it works with an important assumption: an accurate estimation of the SVM boundary. As noticed in the previous sub-section, when labels are too few, this estimation is not trivial. Furthermore, the minimum of labels (20 in [8]) required for a good exploitation of this method is not easy to tune. In experiments, we noticed that this minimum depends on the searched category, and may greatly vary.

3.2. Method for image set selection

We introduce a method with the same principle than SVM$_{active}$, but without using the SVM boundary to find the threshold $s(i)$. Indeed, we notice that, even if the boundary may change a lot during the first iterations, the ranking function $r()$ is quite stable. The efficiency of the set selection method is mostly depending on the $s(i)$ estimation.

Our method is based on a adaptive tuning of s during the feedback steps. We propose to analyze the set of labels provided by the user at the ith iteration in order to determine the next value $s(i+1)$.

Actually, we just suppose that the best threshold s_o corresponds to the searched boundary. Such a threshold s_o 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 $s(i)$ is good. We exploit this property to tune s.

At the ith feedback step, the system is able to classify images using the current training set. The user gives new annotations for images $I_{r(i,s(i))}, \dots I_{r(i,s(i)+m-1)}$, and they are compared to the current classification. If user mostly gives relevant annotations, thus classification seems to be good to the rank $s(i)+m-1$. The system can propose images for labelling from an higher rank to get more irrelevant annotations. On the contrary, if user mostly gives irrelevant annotations, thus classification does not seem to be good to the rank $s(i)+m-1$. The system can propose images for labelling from a lower rank to get more relevant annotations.

Thanks to this approach, we expect the same behavior than SVM$_{active}$, but without problems due to few training data.

3.3. Algorithm

At the beginning, $i=1$, $s(i)=0$, and the system proposes to user the m most relevant images $I_{r(1,1)}, \dots, I_{r(1,m)}$ to label. The user gives m annotation $A_1, \dots, A_m$. The system computes the number $r_{rel}(i)$ of relevant annotations and the number $r_{irr}(i)$ of irrelevant annotations in $A_1, \dots, A_m$. The next rank $s(i+1)$ is computed using the following relation:

$$s(i+1) = s(i) + h(r_{rel}(i), r_{irr}(i)) \qquad (2)$$

where $h(.,.)$ is a function which characterizes the system dynamics. For now, we choose: $h(x,y) = k \times (x-y)$.

Once $s(i+1)$ is computed, the system proposes to the user the m images from $I_{r(i+1,s(i+1))}$ to $I_{r(i+1,s(i+1)+m-1)}$.

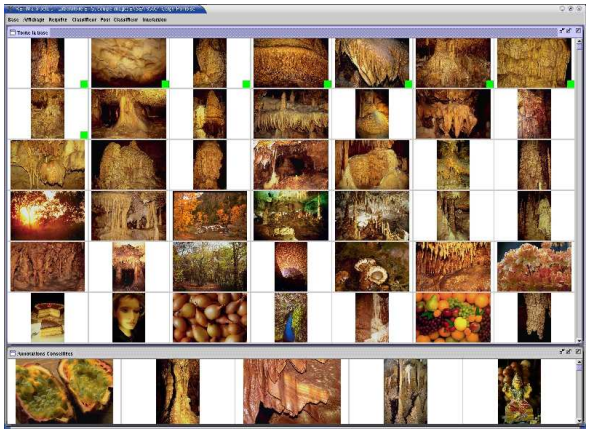

Fig. 3. RETIN AL User Interface

4. TRAINING SET OPTIMIZATION

It could be relevant to take into account how the selected images are correlated between them. For instance, if user labels images close each other in the feature space, the classification should not differ a lot. Only one of these images should be proposed to user. Furthermore, asking the user for labelling an image close to another already labelled is also useless.

To overcome this problem, we propose a method to increase the sparseness of training data. We compute m clusters of images from $I_{r(i,s(i))}$ to $I_{r(i,s(i)+M-1)}$, using an enhanced version of LBG algorithm [9], with $M >> m$. Next, the system selects for labelling the most relevant image in each cluster. Thus, images close each other in the feature space will not be selected together for labelling.

5. EXPERIMENTS

5.1. RETIN AL System

RETIN AL is a new version of the CBIR system developed in ETIS laboratory. User interface is compound of two windows (cf Fig. 3). One displays images in decreasing order of relevance (upper window), and another displays the suggested images to label (lower window). System uses a two-class SVM to classify database, but in the case where only one kind of labels is provided, a one-class SVM is used. In both cases (one-class and two-class), a gaussian kernel with a χ^2 distance is used.

5.2. Evaluation Protocol

Image database used for experiments is an extract of 6,000 images from COREL photo database. Features are the distribution of $CIELab$ color and Gabor Filters. To perform interesting evaluation, we built from this database 11 categories[1] of different sizes and complexities. Some of the categories have jointly images (for instance, castles and mountains of Europe, birds in savannah). For any category search, there is no trivial way to perform a classification between relevant and irrelevant pictures.

The CBIR system performances are measured using precision/recall curves, and the average precision[2].

5.3. Results

RETIN AL Parameters. The number of labelled images m per feedback step and the factor k in function $h(.,.)$ play an important role. We set $k = 2$, and examine different values of m (8,15,30,60) keeping the total of annotations constant (120). Results are displayed in Fig. 4. Globally, as m increases, precision decreases for lower values of recall, and increases for higher values of recall. Supposing that user wants a maximum of precision in the first displayed images, the system has to ask the user to label few images at each iteration.

Active Learning evaluation. We compare our method to SVM$_{active}$ as described by Tong [8]. Table 1 shows the results for the 11 categories. RETIN AL has the best performances for all the categories, followed by SVM$_{active}$ which shares those performances for half of the categories. We can also see on a precision/recall curve (cf. Fig. 5) that RETIN AL has higher precision values for lower values of recall.

Transductive SVM. Transductive SVM needs an adaptation to CBIR context to be comparable to other methods. This method requires the number of vectors to be put in relevant class. In simulations, we set it to the number of images in desired category. We use only TSVM to compute the class of each image, distance to boundary of inductive SVM is used for computing relevance to category. Otherwise, performances can be very low. As curves in Fig. 6 show, transductive SVM does not improve performances for this test category. Actually, we noticed that the transductive approach sometimes improves results, sometimes not. It is very data-dependent, and, of course, time consuming [3].

6. CONCLUSION

In this article, we present an efficient active learning method (RETIN AL) for content-based image retrieval. We introduce an algorithm to select the most difficult images to classify with only few training data. In addition to this technique, we present an approach to select sparse images in the feature space. We also propose an adaptation of Transductive SVM to CBIR context.

The method has been validated through experiments and compared to a reference active learning method. The results show that it is a powerful tool to improve performances.

[1] A description of this database and the 11 categories can be found at: http://www-etis.ensea.fr/∼cord/data/mcorel.tar.gz

[2] cf. TREC VIDEO conference: http://www-nlpir.nist.gov/projects/trecvid/

category	RETIN AL	SVM$_{active}$
birds	**31**	**31**
castles	**38**	**38**
caverns	**78**	75
dogs	**58**	**58**
doors	**93**	83
Europe	**35**	**35**
flowers	**67**	57
food	**71**	59
mountains	**54**	**54**
objects	**78**	76
savannah	**68**	56

Table 1. Performances: average precision in % (initialization with 21 examples, 20 annotations per feedback, 9 feedback steps).

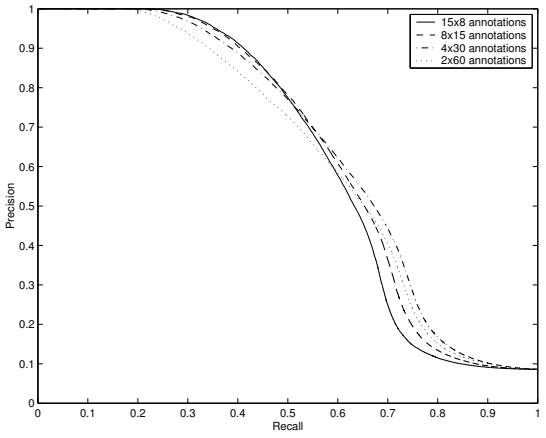

Fig. 4. Influence of parameter m: precision/recall curves for 2,4,8 and 15 feedback steps.

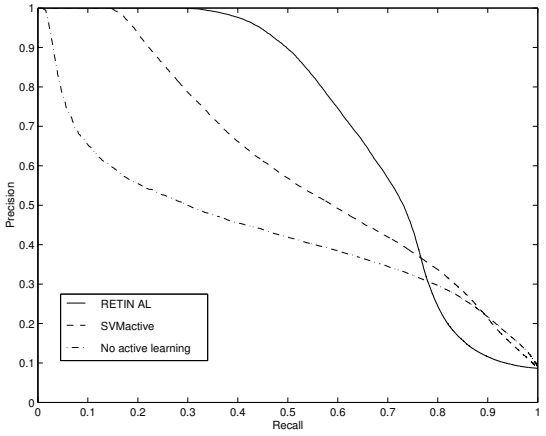

Fig. 5. Precision/recall curve comparison: RETIN AL, SVM$_{active}$ and SVM without active learning.

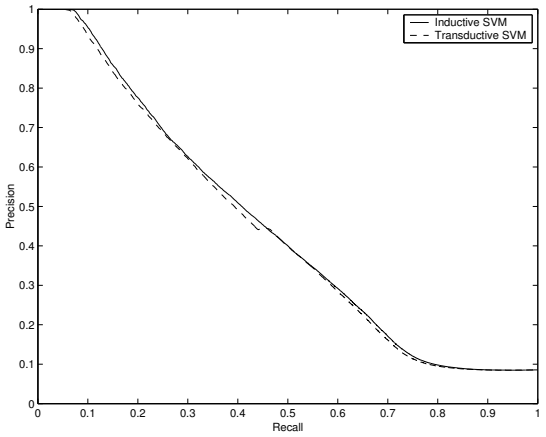

Fig. 6. Comparison between transductive and inductive SVM (training set of 20 relevant and 20 irrelevant images).

7. REFERENCES

[1] R.C. Veltkamp, "Content-based image retrieval system: A survey," Tech. Rep., University of Utrecht, 2000.

[2] N. Vasconcelos, "Content-based image retreival from image databases: current solutions and future directions," in *IEEE International Conference on Image Processing*, Thessaloniki, Greece, october 2001.

[3] E. Chang, B. Li, G. Wu, and K. Goh, "Statistical learning for effective visual information retrival," in *IEEE International Conference on Image Processing*, 2003.

[4] O. Chapelle, P. Haffner, and V. Vapnik, "Svms for histogram based image classification," *IEEE Transactions on Neural Networks*, , no. 9, 1999.

[5] T. Joachims, "A statistical learning model of text classification with support vector machines," in *Proceedings of the Conference on Research and Development in Information Retrieval (SIGIR), ACM*, 2001.

[6] B. Le Saux, *Classification non exclusive et personnalisation par apprentissage : Application à la navigation dans les bases d'images*, Ph.D. thesis, INRIA, 2003.

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

[8] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *ACM Multimedia*, 2001.

[9] G. Patanè and M. Russo, "The enhanced LBG algorithm," *IEEE Transactions on Neural Networks*, vol. 14, no. 9, pp. 1219–1237, November 2001.