RETIN: a Smart Interactive Digital Media Retrieval System

Julien Gony, Matthieu Cord
LIP6, CNRS, Université P.M. Curie
Paris, France

Sylvie Philipp-Foliguet, Philippe H. Gosselin, Frederic Precioso, Michel Jordan
ETIS, ENSEA/UCP/CNRS
Cergy, France

ABSTRACT

This demonstration presents a digital media retrieval system for searching large categories in different media databases. The core of our system is an interactive online classification based on user labeling. The classification is obtained with a statistical learning method: kernels for similarity representation and SVM (Support Vector Machine) using binary user annotations. RETIN applies also an active learning strategy for proposing documents to the user for labeling. The system can deal with images, 3D objects and videos and other media can be added to. A graphical user interface allows easy browsing of different media, simple and user-friendly interaction and fast retrieval.

Categories and Subject Descriptors

H.3.1 [**Information Storage And Retrieval**]: Content Analysis and Indexing—*Indexing methods*; H.3.3 [**Information Storage And Retrieval**]: Information Search and Retrieval—*Retrieval models, Relevance feedback*

General Terms

Algorithms

Keywords

Machine Learning, Content-Based Retrieval, Multimedia

1. INTRODUCTION

Human interactive systems have attracted a lot of research interest in recent years, especially for content-based retrieval systems. Traditional techniques in CBR are limited by the semantic gap, which separates the low-level information extracted from the document and the semantic user request.

To fulfil this semantic gap, interactive systems ask the user to conduct search within the database. Starting with a coarse query, the interactive process allows the user to refine the query as much as necessary. The user interaction consists in binary labels indicating whether or not the document belongs to the desired category (relevance feedback). The main idea of relevance feedback is to use information provided by the user to improve system effectiveness.

In category search, each document has to be classified as belonging or not to the category. The objective of the statistical methods is to update a relevance function or a binary classification of document using the user labels. Discrimination methods (from statistical learning) may significantly improve the effectiveness of visual information retrieval tasks. This approach treats the relevance feedback problem as a supervised learning problem. A binary classifier is learnt by using all relevant and irrelevant labeled images as input training data.

The RETIN search engine, developed in ETIS lab, addresses the problem of category search. Our strategy focus on the online retrieval step to reduce the semantic gap. This is achieved with statistical learning techniques applied on media document category retrieval. RETIN is based on kernels functions to represent similarity, Support Vector Machines for classification and on an active learning strategy to boost the training step. See [2],[6],[5] for more informations.

2. RETIN SYSTEM OVERVIEW

RETIN is composed of two parts. The first one is the offline indexing of the database, where we compute different descriptors for each media, and store the results in an index database. The second one is the online search which is composed of a graphical user interface for query, results display and user labeling, and a search engine using an active learning strategy method. RETIN scheme is shown on figure 1.

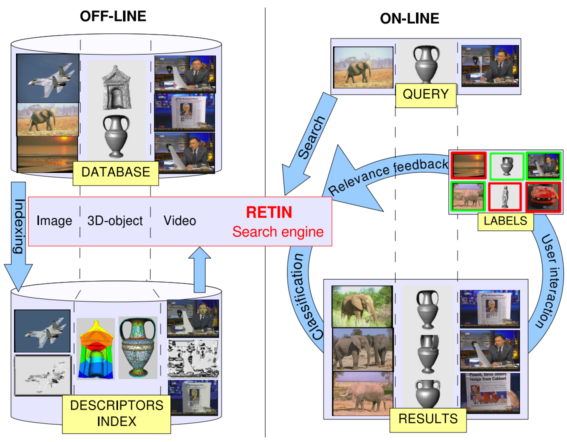

Figure 1: RETIN search engine scheme.

2.1 RETIN Interactive retrieval

The system uses an active learning scheme based on binary classification in order to interact with a user looking

CIVR'07, July 9–11, 2007, Amsterdam, The Netherlands.
ACM 978-1-59593-733-9/07/0007.

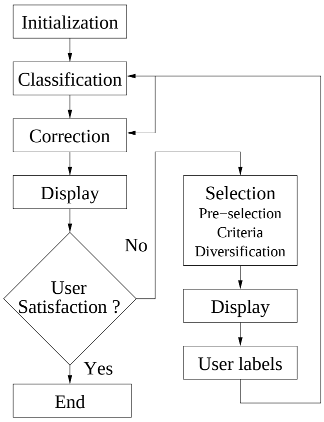

Figure 2: RETIN active learning scheme.

for documents concepts in databases. The scheme is summarized in figure 2. More details can be found in [2].

Initialization

A retrieval session is initialized from one document brought by the user. The features are computed on that new document and added to the database. This document is then labeled as relevant, and the closest documents are shown to the user.

Classification

A binary classifier is trained with the labels the user has given. We use a SVM with a Gaussian χ^2 kernel, as it has revealed being the most efficient [5]. The result is a function $f_{\mathcal{A}_{\mathbf{y}}}(\mathbf{x}_i)$ which returns the relevance of each document $\mathbf{x}_i$, according to the examples $\mathcal{A}_{\mathbf{y}}$.

Active selection

We add an active correction to the boundary (border between relevant and irrelevant) in order to deal with the few training data and the imbalance of the classes. Since our aim here is to enhance the selection, we propose to consider the boundary as the area of uncertainty, because of the interest in our context of uncertainty-based active learners in terms of efficiency and complexity. The system selects a set of documents the user should label. The selection may be such as the labeling of those documents will enhance the most the classification. We divide the selection in three steps. First, we reduce the computational time, by pre-selecting around five hundreds documents which may be in the optimal selection set (closest documents to the corrected boundary). Then we compute the selection criteria (maximizing the Average Precision) for each pre-selected documents. Finally, we compute the batch selection using the criteria.

Feedback

The user labels the selected documents, and a new classification and correction can be computed. The process is repeated as many times as necessary.

2.2 Supported media

RETIN was initially developed for CBIR context, but other medias have been integrated into the system. Videos, since it can be seen as sequence of images (video is divided into shots, key-frames are extracted for each shot and CBIR method are used for retrieval). 3D-objects retrieval is also integrated, the features extracted from the 3D-models are based on shape.

2.3 Indexing

Each document of the database is represented by features which aim at giving a complete and compact description of the content of the document, easy to compute and to deal with. These features can be used in the RETIN search engine since they are represented by a vector of the same size for one feature for each document of the database.

2.3.1 Images

Color and texture information are exploited. As none of color spaces has proved its superiority over the others for image coding, the HSV space is used by default. For texture analysis, Gabor filters are used with twelve different scales and orientations. Signatures are statistical distributions of colors and textures resulting from a dynamic quantization of the feature spaces. That means that we use color and texture space clustering to compute the image histograms. Both spaces are clustered using an enhanced LBG algorithm.

For a generalist database (around 10, 000 images), a small number of classes obtained by a dynamic clustering of the database is sufficient to build efficient signatures. We have adopted this dynamic quantization in RETIN with 25 classes (default value). The reader can find more details in [6].

2.3.2 3D objects

Several descriptors are available in RETIN for indexing: cord histograms, extended Gaussian images (EGI) and 3D Hough transform.

Cord histograms: A cord is defined as the vector from the object center to a vertex. Three features are defined from the cords: the length of the cord and the two angles between the cord and the first principal axis, and the second principal axis.

EGI: Each 3D object is projected onto a Gaussian sphere, and each point of the sphere is valued with the area of the object faces of the same orientation. CEGI and 3D Hough are variant of EGI, all these descriptors are detailed in [4].

2.3.3 Videos

The video indexing is performed in two parts. The first one is the segmentation of the video into shots, the second one is the computing of the features for each shot. Each shot can then be summarized by one or several key-frames.

The shot extraction is composed of two boundary detection algorithm. The first one is for detecting abrupt transitions, and the second one is for gradual transitions. To detect abrupt transitions, feature vectors with frame-to-frame feature differences is computed and a kernel-based SVM classifier is used to separate cuts and non cuts after training. Once the video sequence is segmented into cut-free segments, we process a three-level gradual transition detection. The reader can find more details in [1].

Key-frame extraction is based on a clustering of each segmented shot. For each cluster, the closest frame to the cluster center is taken as a key-frame. The features extracted for the shots are L*a*b color and Gabor texture Retin features for still images (key-frames) and the Fourier-Mellin and Zernike moments extracted for shot detection.

2.4 Graphical User Interface (GUI)

The user interface is divided in two panels (Fig. 3 (top)). On top window, documents in decreasing order of relevance are displayed, according to the current classifier. On bottom window, documents proposed for labeling are displayed, according to the current active learner. The user is invited to follow advises in bottom window, but he can choose to label documents in the top window. GUI for 3D-object and video retrieval is shown in Fig 3 (bottom).

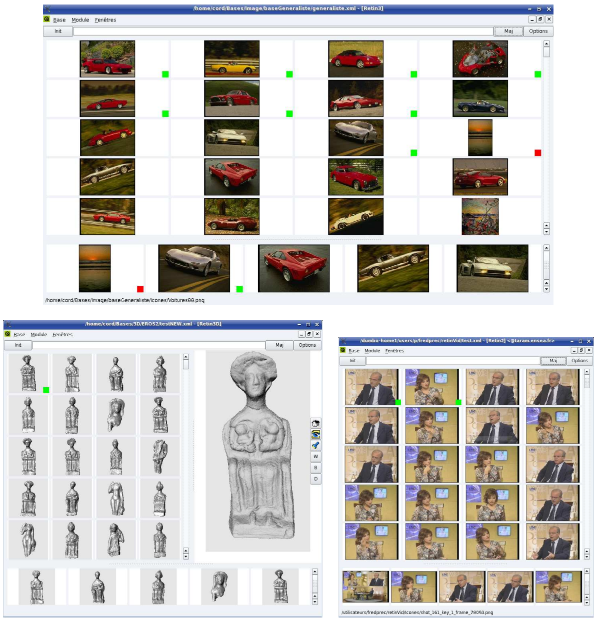

Figure 3: RETIN GUI: images (top), 3D (left), Videos (right).

3. DEMONSTRATION

We give here an example of a RETIN search session on image database. As it is the same type of interaction for 3D and video modules, it can be generalized to all these database retrievals. In this example, the user is looking for images that represents or contains a plane.

The first step of a session is to choose an image that is belonging to the category that the user is searching. The user selects an image with a plane (the image with the green square on fig. 4).

Then, a first retrieval result is displayed (fig. 5). The classification with only one labeled image has retrieved some planes. To improve results, the user labels the 5 proposed images (all of these are relevant).

A new classification is computed with this five new labels and returns a new result (fig. 6). Then a new labeling operation is done. This time, three relevant images and two irrelevant images have been proposed.

After two other iterations, 21 images have been labelled (15 relevant, and 6 irrelevant). We must go down to the 4th page of the classified results (fig. 7) to find the first irrelevant image (all the 67 first images are relevant).

This demonstration will be proposed during the conference for different media databases (3D objects, videos).

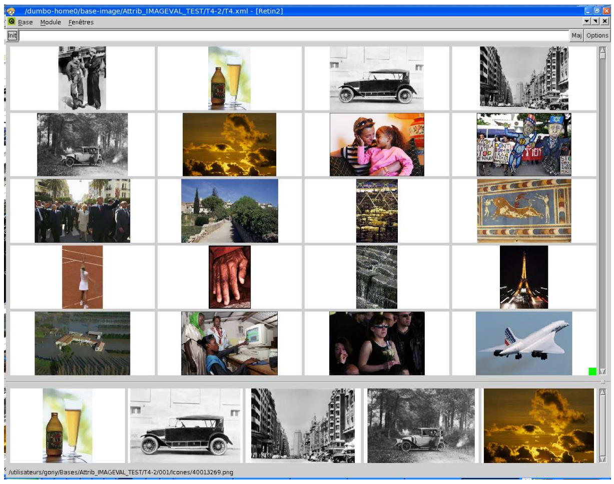

Figure 4: Initialization of the query.

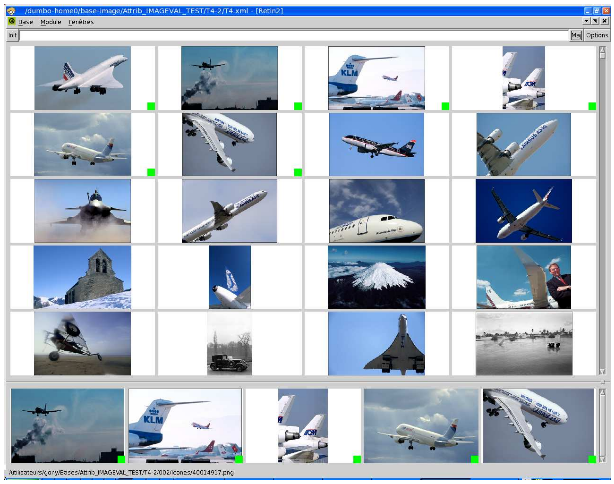

Figure 5: First results and first labels.

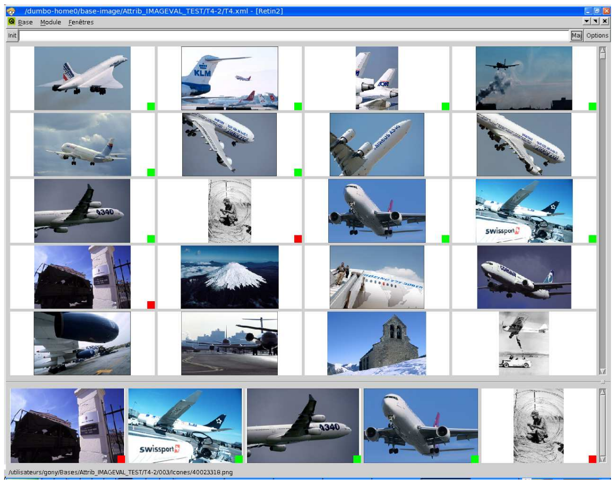

Figure 6: Second results and second labels.

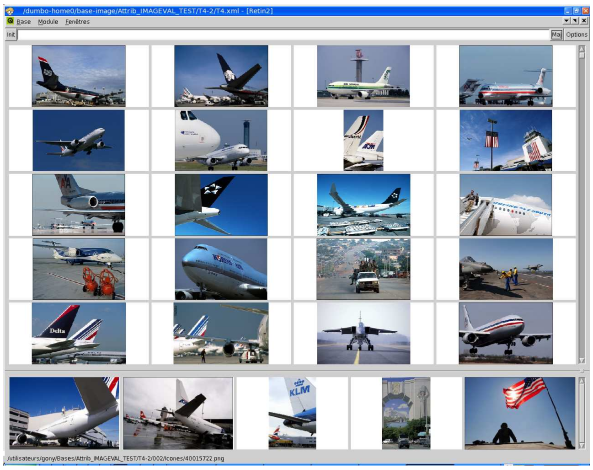

Figure 7: Results of the 4th page.

4. EXTENSIONS

4.1 Local search module

FReBIR (Fuzzy Region Based Image Retrieval) is a region-based image retrieval system, in which images are represented as adjacency graphs of fuzzy regions. The goal of this retrieval module is to retrieve images containing a specific object. An algorithm to match sub-graphs of fuzzy regions is then applied in order to retrieve images from partial queries, taking into account the image composition. Images are represented by a set of fuzzy regions, with their features (the same as global RETIN engine for images) and the composition of the image is stored into an attributed relational graph (ARG) of regions, aiming at representing the relative positions of regions. Then the problem of image retrieval from a partial query can be seen as a problem of inexact graph matching of ARGs. RETIN is used here to compute the dissimilarity between regions thanks to user labels on images. More informations about FReBIR are in [7]. This module will be also available for demo during the conference.

Figure 8: FReBIR: Query Regions (top), Results (bottom).

4.2 Other modules based on RETIN

The long term learning module of RETIN uses all the labels accumulated during many interactive uses of any retrieval system to improve the feature representation of the documents. With such an optimized representation, we attempt to get a better match with semantic concepts. See [3] for more details.

In order to manage large collections of images distributed on networks, a new architecture for media retrieval in distributed media databases is currently developed, based on multi-agent systems. Our system, inspired by "ant algorithms", uses labels provided by the user for learning both the searched category of documents and the path to the most relevant databases. The problem of interactive retrieval into distributed databases context is considered with a decentralize approach. It is based on Multi-Agent Systems (MAS) offering interesting properties in comparison with centralized systems : since the agents are distributed, several sub-tasks can be processed in parallel, thus sparing both CPU and bandwidth. If a machine is down, machines still up can keep on processing the search task. More details are in [8]

M. Cord recently joined the LIP6 laboratory, and an extension of RETIN to text/image multimedia search engine is also currently under development.

4.3 Web interface

A web graphical interface of RETIN is already available at `http://dupont.ensea.fr/~ruven/new.php`. This version allows for the moment image retrieval in a generalist database. New application integration, such as 3D, videos and local search (FReBIR) is on the roadmap.

Aknowledgement

The authors are grateful to MUSCLE Network of Excellence for the financial support of this work.

5. REFERENCES

[1] G. Camara-Chavez, M. Cord, S. Philipp-Foliguet, F. Precioso, and A. de Albuquerque Araujo. Robust scene cut detection by supervised learning. In *EUSIPCO*, Firenze, Italy, 2006.

[2] M. Cord, P. Gosselin, and S. Philipp-Foliguet. Stochastic exploration and active learning for image retrieval. *Image and Vision Computing*, 25:14–23, 2007.

[3] M. Cord and P.-H. Gosselin. Image retrieval using long-term semantic learning. In *IEEE International Conference on Image Processing*, Atlanta, USA, oct. 2006.

[4] D. Gorisse, M. Cord, M. Jordan, and S. Philipp-Foliguet. 3d content-based retrieval in artwork databases. In *Proceedings of the 3DTV-Confence*, Kos Island, Greece, May 2007.

[5] P. Gosselin and M. Cord. Feature based approach to semi-supervised similarity learning. *Pattern Recognition*, (39):1839–1851, 2006. Special issue: Similarity-Based Pattern Recognition.

[6] P.-H. Gosselin and M. Cord. A comparison of active classification methods for content-based image retrieval. In *Int. Workshop on Computer Vision Meets Database, CVDB*, Paris, juin 2004.

[7] S. Philipp-Foliguet and J. Gony. Frebir : Fuzzy region-based image retrieval. In *IPMU*, Paris, France, july 2006.

[8] D. Picard, M. Cord, and A. Revel. Cbir in distributed databases using a multi-agent system. In *IEEE International Conference on Image Processing (ICIP'06)*, 2006.