IMAGE COLLECTION STRUCTURING BASED ON EVIDENTIAL ACTIVE LEARNER

Hervé Goëau, Olivier Buisson, Marie-Luce Viaud

Institut National de l'Audiovisuel
4, avenue de l'Europe
94366 Bry-sur-marne Cedex, France
{hgoeau, obuisson, mlviaud}@ina.fr

ABSTRACT

Organising a collection of images requires an intensive and time consuming human effort. We present here a framework to classify dynamically collections of images without a priori content knowledge. Our work is based on active learning techniques: unlabeled samples are selected iteratively one by one, and a knn-evidential classifier make a proposition of label at each step. Users can initialize, remove or merge classes and may correct the propositions. The Transferable Belief Model framework offers us a complete formal model to express jointly the classifier and different sampling strategies such as positivity, ambiguity and diversity. Our aims are to study these different sampling strategies in order to minimize the error rates as well as the user cognitive charge according to the distribution of the endeavor over time.

1. INTRODUCTION

The French National Audiovisual Institute (INA) missions are to preserve the French audiovisual heritage, insure its exploitation and make it more readily available. To achieve these objectives, archivists identify and annotate day by day a sample of the French TV programs over a set of more than 100 broadcasted channels. Then, they structure and organize by themes and collections the wholeness of INA's corpora, which represents about 2M of hours of TV programs and 100,000 hours of radio. But, INA owns also more than 2M of pictures and around four hundred thousands of them have been already scanned for an upcoming human annotation. These pictures have few contextual information (who, when, where), and no fine date-stamping. As a consequence, it is not possible to make any temporal clustering like in common system of picture management. Moreover, archivists may not have any knowledge of their content before starting the annotation. The work presented in this paper is focused on the creation of a user supervised tool to help archivists in the pre-annotation task, i.e. the creation of groups (or classes) of pictures for semantic labeling.

This kind of tool deals with the well-known problem of "semantic gap". Indeed, the similarity between the pictures perceived by humans does not necessarily match the similarity computed in a feature description space. To make things worse, semantic concepts are specific to each user and may even differ according to users' mood.

Actually, numerous methods attempt to reduce the semantic gap by making statistical or probabilistic associations of low level visual features to high level concepts. But these technologies may not be ready yet to match industrial constraints at INA for several reasons. First, the number of concepts used for the annotation is around 20,000 while the maximum number of concepts computable is around 1000. Moreover, computable concepts do usually not match concepts used by archivists for annotating documents. And because the annotation is manual, the quality required is very high: no error is accepted during the process. To respond to this constraint of high accuracy, our system uses principles of active learning in relevant feedback (RF): users may modify the proposition of the system at any time and then the classification process adapts dynamically.

We choose to formulate our problem in the Transferable Belief Model (TBM) because this formalisation lets us take into account imprecision, uncertainty and conflict inherent to the visual features description. Within our framework, we design an image classifier which formulates a distance rejection by all classes and potential labeling in a class. Moreover, this evidential active learner lets us build several sampling strategies. We propose to use multiple criteria: positivity, ambiguity and diversity to select the photos presented to the user at each step.

In the first part of this paper, we describe our evidential framework of classification with the TBM formalism. Then, we present the graphic user interface, and experiments on different sampling strategies. Finally, we analyze and comment our work and conclude on the perspectives offered by our framework.

2. APPLICATION CONSTRAINTS

At the beiginning of a work session, an user will generally have no idea about the images content, the number and the the kind of classes he could define. He probably will hesitate and make some mistakes or correct a previous decisions. Indeed, the concept of each class is refined at each new labeled image and it is often at the end of the process, when all images have been seen, that the choice of the classes may be confirmed in the user mind. Then, the system must be readily adaptable and must offer all the classical editing functionalities:

- user can create, suppress, split and merge the labeled classes as often as he wishes,
- each image should be easily moved in another class, allowing the user to change his mind at any time during the process of classification

The system should take into account instantaneously any user modification. Moreover, the system must offer ways to give a low cognitive charge to the user, in order to minimize his effort. In this direction, the system must be able to bring out several suggestions of labeled classes:

- focused on one class, if the user wants to find all the potential images in,
- or focused on several classes if he wants to disambiguate them.

Finally, a complete image classification system must deal with the distance rejection and the diversity of visual content.

- the system must be able to detect and propose potential new classes if the content is very dissimilar with all the labeled classes,
- moreover, it must be possible to put eventually these dissimilar images in an existing labeled class, and as a consequence the classifier should be able to deal with multi prototype classification.

These two last points are main issues because the distance rejection is not so often taking into account, and multiclass classification is a real challenge. In this first approach, we suppose that each image belongs to an unique class, corresponding to its "dominant meaning" from the user point of view.

Moreover, we will show further that this information of visual diversity may be very useful in strategy sampling.

3. THEORITICAL ISSUES

Unsupervised methods like in [1] aggregate visual descriptors in the associated feature space. But it's not certain that the clusters map the semantic classes generated manually by an user. If the user wants to create classes with heterogeneous visual content, this method is not well appropriate.

With automatic process, the quality of the created clusters can be very disappointing, and the system may be rejected by the user. Some semi-supervised clustering methods allow the user to partially control the system. For example in [2], [3] the system asks the user to validate or not couples of pictures. But, if this method is well suited to mono prototype classes, it may not be relevant for composite classes.

The work describes in [4] is very close to our targets. In this method, the classes or clusters are initialized by a hierarchical clustering algorithm. The user can optimize the classification by a multiclass classifier based on SVM. But in such methods, the number of classes is given, making difficult the dynamical creation or suppression of classes. Generally, for converging to an acceptable class arrangement, it is required to consider a close world assumption. We think that it may be interesting to consider the open word assumption because it allows the system to accept a class of "unclassified" images, as the user may do in a manual process.

During the process, we would like that the action of the user are restricted to move images or classes. Any parameters or thresholds will be hidden from the user who should only concentrate on the semantic of his classification

To solve this iterative (interactive) process of classification according to all these constraints, we have chosen active learning methods. Active learning systems receive a lot of interest from academic and industry because they offer a promising solution to the semantic gap problem. In fact, contrary to the early systems, focused on fully automatic strategies, these approaches allow the introduction of human computer interaction into Content-Based Image Retrieval (CBIR) like in [5], [6]. For instance, due to its good generalization ability, relevance feedback (RF) based classification on support vector machine (SVM) learning methods has become popular to improve retrieval performance in CBIR systems using feature representation. Ideally, the system feedback should provide most useful samples for the system and a user has to give binary labels indicating whether or not the photo belongs to the query concept.

Principles are similar in our case. For each image, the system will give a suggestion of classification into an existing labeled class, and the user decides to invalidate it or not. The aim is to limit effort of the user by selecting valuable samples. Usually, active learning is based on a combination of a two modules.

- A learning level to train a classifier on labeled samples.
- A sampling level to select samples for users to label before passing it to the learning system for next iteration.

Therefore, make up an active learning system represents a double challenge. The learning engine is crucial for achieving good classification performance with limited training data, and the sampling engine must choose the most valuable samples for converging to satisfying results. Selecting the most valuable samples at each step of classification can be called "sampling strategy". In the state of the art, different strategies are defined for choosing a sample in [7], [8].

Most Positive (or error reduction strategy): this strategy chooses firstly the samples which have the highest value for labeling in the sampling engine.

Most Informative (or batch-simple strategy) aims at selecting unlabeled data that will give most information to the current classifier. Usually, the criterion corresponds to the **Most Ambiguous**, i.e. a sample between two or several classes in the feature space. For example, in [9] an informativeness-based selection criterion is proposed for a SVM classification. The basic idea is to select the most informative candidates whose representations in the feature space induced by the kernel are closest to the SVM hyperplane.

Diversity encourages the selection of unlabeled samples that are far from all the labeled classes. This criterion removes the redundancy. For instance, in [10] the redundancy of samples is measured by the angles between the samples. This strategy can be useful for discovering new classes or for the definition of new prototypes in multiclass.

In the state of the art, these sampling criterions are rarely proposed together. Ones prefer the positive reinforcement, others the ambiguity. In the next part, we lean on the formal framework of Transferable Belief Model to build an evidential classifier and express some of strategies above-cited together with unusual ones in sampling strategies.

4. EVIDENTIAL ACTIVE LEARNER

4.1. Why an evidential one?

A main objective of an active learning system is to minimize the number of selected samples used to optimize the classifier. In our case, we would like minimizing the number of manual corrections in order to reduce the effort asked to the user. Our aim is to reduce the final error rate but also to be time effective as well: an user's correction is an action while a validation is not expressed. No change in the process is a validation.

An usual choice in active learning is to take a SVM classifier. But it's difficult to express the distance rejection with SVM. Moreover, usually the classifier and the sampling strategy are conceived independently in an active learning system. We prefer to express the classifier and the sampling engine in an unified framework. We choose the Transferable Belief Model, an elaboration on the Dempster-Shafer theory of evidence because:

- it's a formal framework fusion process based on firm ground,
- it enables an intuitive modeling of the knowledge,

- it respects an open-world assumption, i.e. an event not described in an initial frame of reference,
- adding new information is readily possible.

From now, the system deals only with visual features without metadata, but later on, complementary information like the date-stamping in EXIF metadata or face detection may be added to complete the knowledge.

The classifier presented below is inspired from the Knn-Evidential (Knn-Ev) classifier proposed by Denoeux [11].

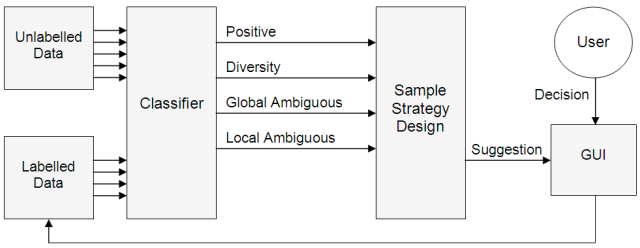

Fig. 1. The framework: the classifier computes different measures on the unlabeled images in accordance with the previous labeled ones. These measures allow to select the most representative of the current strategy, for example the Most Positive or the Most Rejected or the Most Ambiguous. User must validate or not the suggestion of label.

4.2. Formulation of the problem

Let's considering an initial collection $I^0 = \{i_1^0, i_2^0, \cdots, i_N^0\}$ of N images to organise. Each image is associated with visual descriptors and let $X^0 = \{x_1^0, x_2^0, \cdots x_N^0\}$ be the set of corresponding vector descriptors.

At a current step of classification t, a set of Q classes $C = \{C_1, C_2, \cdots, C_Q\}$ has been initialized by a user, and each one contains at least one labeled image. Then, we have $I^t = \{i_{r1}^t, i_{r2}^t, \cdots, i_R^t\}$ the resting set of R images to classify and $I_l^t = \{i_{l1}^t, i_{l2}^t, \cdots, i_L^t\}$ the set of the previously labeled images with their respective descriptors X_l^t.

At each step of classification, the classifier has to make a suggestion for an image i. The process involves distinct stages.

- A local fusion is processed for each class C_q of C in an independently way. The aims is to obtain a local fusion of observations of the k nearest neighbors of i in the feature space.
- A global process takes into account all classes.

4.2.1. Local fusion of k nearest neighbors for one class

Let's consider one class C_q containing some labeled images i_q^t classified previously. Classifying an image i of I^t in a class C_q is described by two states gathered in a frame of discernment:

$$\Omega_q = \{H_q, \overline{H_q}\} \tag{1}$$

with H_q (resp. $\overline{H_q}$) the hypothesis "i is a member of C_q" (resp. "i does not belong to C_q"). A basic belief assignment (BBA) is defined on a set of propositions:

$$2^{\Omega_q} = \{\emptyset, H_q, \overline{H_q}, (H_q, \overline{H_q})\} \tag{2}$$

The set $(H_q, \overline{H_q})$ explicitly represents the doubt concerning the real state of the belonging to a class, and the emptyset $\emptyset$ symbolize a eventual conflict between 2 information sources. Voluntarily, only two propositions are taken into account, the belonging and the rejection:

$$m(\emptyset|x) = m((H_q, \overline{H_q})|x) = 0 \tag{3}$$

$$m(H_q|x) + m(\overline{H_q}|x) = 1 \tag{4}$$

Here, we have distinctions with the initial version of the Knn-Ev classifier. Firstly, the fusion process is without normalization in respect to the open-world assumption. Secondly, eq. 4 define a bayesian BBA, i.e. all its focal elements are singletons [12]. The last difference is a local adaptation of masses at the borders of classes (see section 4.2.2).

Let x_q^j be a descriptor of a labeled image in X_l^t. Thus, it's potentially a nearest neighbor. We define:

$$m^{x_q^j}(\overline{H_q}|x) = 1 - \alpha_q^j(x) \tag{5}$$

$$m^{x_q^j}(H_q|x) = 1 - m^{x_q^j}(\overline{H_q}|x) \tag{6}$$

with a kernel function $\alpha_q^j(x)$:

$$\alpha_q^i(x) = \alpha_0 . e^{-(d(x,x_q^j)/\sigma)^\beta} \tag{7}$$

where $d(x, x_q^j)$ is a distance in the features space between the labeled element x_q^j and x the image to classify. α_0 is a weakening factor arbitrarily fixed at 0.95 such as β which has been fixed to a small value ($\beta = 2$). The radius parameter σ involves a level of confidence around a labeled image.

By combining "2-nn", we have:

$$m^{x_q^j, x_q^k}(\overline{H_q}|x) = (1 - \alpha_q^j(x))(1 - \alpha_q^k(x)) \tag{8}$$

$$m^{x_q^j, x_q^k}(H_q|x) = 1 - (1 - \alpha_q^j(x))(1 - \alpha_q^k(x)) \tag{9}$$

The aim of this formulation is to obtain the OR logical operator. If x and x_q^j are close, but x et x_q^k are farther, $(1 - \alpha_q^k(x))$ will tend toward 1 and $m^{x_q^j, x_q^k}(H_q|x)$ will tend toward $(1 - \alpha_j^q(x))$. In other words, if the observations are very close to x, but only one is far, an important masse on the hypothesis of belonging is preserved.

By considering a set of k nearest neighbor, the result of the combination gives us a new BBA corresponding to the local fusion for one class C_q:

$$m(\overline{H_q}|x) = \prod_{j=0}^{k} (1 - \alpha_q^j(x)) \tag{10}$$

$$m(H_q|x) = 1 - \prod_{j=0}^{k} (1 - \alpha_q^j(x)) \tag{11}$$

4.2.2. Sigma estimation

We choose to maximize the ambiguity generated by a image in order to accelerate the generalization process of the classifier. The aim is that each prototype covers a maximum feature space.

A solution consists to have a local adaptation at the borders of the classes. Thus, an individual σ_q^j is estimated by labeled sample. When a labeled image x_q^j is on a border of a class in the feature space, its σ_q^j tends to be small. As a consequence, if x_q^j is selected as a k nearest neighbor, the σ_q^j will weaken its influence. By contrast,

a central labeled sample of a monoprototype class will have a long σ_q^j which reach the border of the nearest neighbor class. Figure 2 illustrates how each σ_q^j estimated is adapted to the local context and the class proximity. In this artificial example, the size of the sample is proportional to the estimated σ. When two classes are very close like the two ones on the up left corner, the local parameters σ tends to be very small, and by comparison, the isolated classes like the one on the bottom right corner tends to have all long local parameter σ.

The following method is used to estimate the σ of a current labeled sample x_j^q owned by one class C_q.

1. Find the nearest labeled sample neighbor x_k^r among the other R classes $R = C \backslash C_q$. This x_k^r is the most ambiguous sample of x_j^q and the corresponding distance $d_{min} = d(x_j^q, x_k^r)$ is kept.
2. A parameter f controls the maximal ambiguity level between classes. The local σ_j^q is estimated (eq. 12) according to the initial kernel definition (eq. 7).

$$\sigma_j^q = \frac{d_{min}}{2\sqrt[\beta]{\log \alpha_0 - \log f}} \tag{12}$$

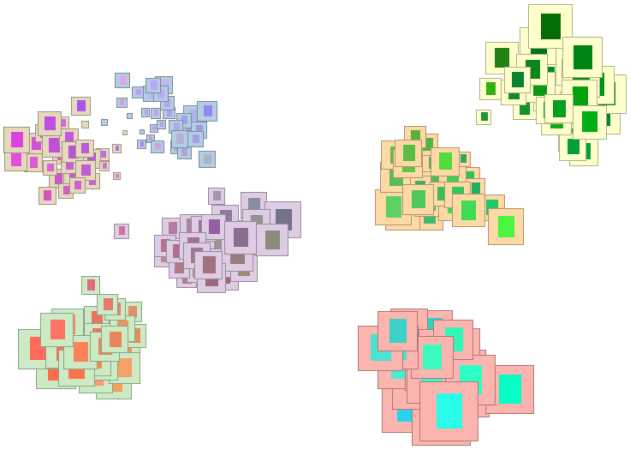

Fig. 2. Local σ estimation on artificial dataset. 7 classes of samples are generated in the RGB color space with a multinormal distribution. The size of images varies proportionately with its local σ estimated.

4.2.3. *Global process for all classes*

The BBAs for each class C_q must be now considered in the same frame of discernment $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_Q$. The Vacuous Extension operator [13] allows to combine the BBAs of each subframe of discernment Ω_q, and gives one BBA containing in our case 2^Q masses. Each mass represents a Q-tuple ω corresponding to one combination of the basic hypothesis A_q (i.e. H_q or $\overline{H_q}$) in each subframe of discernment Ω_q.

$$m^\Omega(\omega|x) = m^\Omega((A_1, A_2, \cdots, A_Q)|x) = \prod m^{\Omega_q}(A_q|x) \tag{13}$$

The masses can be brought together in four categories describing four kinds of propositions for an unlabeled image. The first case, is when the belonging is located on one class C_q and a rejection is indicated by all the others $\{C_{r_1}, C_{r_2}, \cdots, C_R\}$, with $R = C \backslash C_q$. We will identify them as positive propositions because they can be used for labeling images. There are Q masses associated to this kind of proposition:

$$m^\Omega((H_q, \overline{H_{r_1}}, \overline{H_{r_2}}, \cdots, \overline{H_R})|x) = m^{\Omega_q}(H_q|x) \prod_{r \neq q} m^{\Omega_r}(\overline{H_r}|x) \tag{14}$$

Others masses represent conflict informations. In our context of exclusive classification, only one class is possible for an image. As a consequence, if some of the corresponding masses are different from zero, these cases highlight a deficiency in the system. By disambiguating these conflicting samples, the classifier should improve the quality of classification.

One mass represents a global conflict:

$$m^\Omega((H_1, H_2, \cdots, H_Q)|x) = \prod_q m^{\Omega_q}(H_q|x) \tag{15}$$

The conflict can be more selective. Indeed, with our formulation $2^Q - (2+Q)$ masses are available and describe all the local conflict cases between P classes $\{C_{p_1}, C_{p_2}, \cdots\}$ of $\{P\}$ a subset of the Q classes with $2 \leq card(P) < card(Q)$, and $R = \{C_{r_1}, C_{r_2}, \cdots\}$ the resting subset $R = C \backslash P$.

$$m^\Omega((H_{p_1}, H_{p_2}, \cdots, H_P, \overline{H_{r_1}}, \overline{H_{r_2}}, \cdots, \overline{H_R})|x) = \prod_{p \in P} m^{\Omega_p}(H_p|x) \prod_{r \notin P} m^{\Omega_r}(\overline{H_r}|x) \tag{16}$$

A last situation corresponds at the case of the distance rejection which means that the image is too far from all the classes, and a no label may be proposed.

$$m^\Omega((\overline{H_1}, \overline{H_2}, \cdots, \overline{H_Q})|x) = \prod_q m^{\Omega_q}(\overline{H_q}|x) \tag{17}$$

4.2.4. *Decisions and propositions*

In our interactive system, the classifier makes a proposition of class for each image and the user decides. For making a proposition, the pignistic transformation ([14]) is used which consists to put in a equiprobability way the mass of one proposition B of Ω on all hypothesis contained in B. The pignistic probability is defined by:

$$BetP\{m^\Omega\}(\omega|x) = \frac{1}{1 - m^\Omega(\emptyset|x)} \sum_{B \subseteq \Omega, \omega \in B} \frac{m^\Omega(B|x)}{|B|} \tag{18}$$

where a ω is one of the proposition in Ω (one of the Q-tuple defined in previous section 4.2.3). As a consequence, the classifier makes a proposition ω_d by taking the maximum of the pignistic probabilities.

$$\omega_d = \arg\max_{w_i \in \Omega} BetP\{m^\Omega\}(w_i|x) \tag{19}$$

Concretely, ω_d may indicate a distance rejection, a conflict case or a positive case of labelisation in a class. However, if a user wants absolutely a proposition of labelisation, it's possible to take the maximum of pignistic probability of the Q positive expressions (eq. 14):

$$\omega_d' = \arg\max_{w_i \in \Omega_P} BetP\{m^\Omega\}(w_i|x) \tag{20}$$

with Ω_P a sub part of the frame of discernment Ω containing the all positive Q-tuples.

4.3. Sampling Strategies

Images are validated one by one, so there are as many steps as images to classify. At each step, according to a preselected sampling strategy, the classifier chooses one of the unlabeled samples, and makes a suggestion of labeling in one class. According to section 3, different sampling strategies are defined here, directly from the output of the evidential classifier. The Most Positive (MP) (resp. Less Positive (LP)) is the sample x which has the highest (resp. lowest) maximum of pignistic probability on the hypothesis of a belonging of only one class:

$$BetP_{max}(x) = \max_{w_p \in \Omega_P} BetP\{m^{\Omega}\}(w_p|x)$$
$$MP(X^t) = \arg\max_{x \in X^t} BetP_{max}(x) \quad (21)$$

with ω_p a positive proposition in Ω_P (eq. 14). This strategy is useful for labeling very similar samples, by selecting the "easiest" samples. It is also the most intuitive approach for human use. Users may prefer that the system shows good propositions at the beginning of the process, because he doesn't have yet a precise idea of the content of the classes.

In the opposite way, treating the Less Positive first maximises the error risk. This strategy can be useful for generalising faster. The work is more difficult at the beginning of the process, but may be easier at the end.

A second family of strategy is based on Most Conflicting (resp. Less Conflicting) measures. The selected sample x is the one which has the highest (resp. lowest) maximum of pignistic probability on cases corresponding to a conflict between all or several classes (eq. 15 and eq. 16). Thanks to the TBM formalism all conflicting cases can be expressed, from the most global to the most selective conflicting. The Most Global Conflicting is expressed by:

$$MGC(X^t) = \arg\max_{x \in X^t} (\omega_{gc}|x) \quad (22)$$

with ω_{gc} the Global Conflicting Q-tuples (eq. 15). The MGC image means to the user that the image can be potentially labeled in one of the available classes, but the system doesn't know in which one. In sampling strategy, this case corresponds to a case of **global amgibuity**.

The Most Local Conflicting (resp. Less Local Conflicting) is expressed by:

$$BetP_{max}(x) = \max_{w_{lc} \in \Omega_{LC}} BetP\{m^{\Omega}\}(w_{lc}|x)$$
$$MLC(X^t) = \arg\max_{x \in X^t} BetP_{max}(x) \quad (23)$$

with ω_{lc} one of the proposition in Ω_{LC} a sub part of the frame of discernment Ω containing the all Local Conflicting Q-tuples (eq. 16). This kind of strategy is useful for disambiguating the border between the selected classes. Even this strategy gives hard work to the user, it may bring up an improvement at the end of the classification. In sampling strategy, this case corresponds to a case of **selective amgibuity**.

The Most Rejected (MR) is the sample x which has the highest maximum of pignistic probability of all hypothesis:

$$MR(X^t) = \arg\max_{x \in X^t} (\omega_r|x) \quad (24)$$

with ω_r the distance rejected Q-tuples (eq. 17). The MR strategy can be useful for looking **diversity** in terms of visual content. In fact, this strategy responds partially to the "zero page" in the first steps of a work session. At the beginning, when no labeled class has yet be defined, the system chooses randomly one image. The MR strategy will propose the most dissimilar ones. Then, the user can decide to create a new class or to put it in a already defined class. He can repeat this strategy to get an overview of the most dissimilar content.

5. EXPERIMENTATIONS

5.1. The Graphic User Interface

For exploiting the capacity of the classifier, a Graphic User Interface (GUI) has been designed to allow the user to interact with the system. Unlabeled images appear on the vertical heap on the left of the screen (see figure 3). Each class is represented by a horizontal list of images on the right of the screen. The first image on the top of the unlabeled list is the sample selected by the current strategy. At the beginning of the process, the top left image moves out of the heap to reach the center left of the screen and upscale for improving the user observation and analysis of its visual content. Then, it moves smoothly from the left to the right towards the horizontal list of the chosen class and integrate it. More positive is the image for the classifier, shorter the scrutinizing time will be.

Colors are used in the front of each labeled lists to indicate to the user the state of the classifier according to the current image: red for the classes which reject the image, yellow if there is a conflict and green if the class is chosen. The user may leave the process, or drag and drop any image from a list to another if he doesn't agree with the current proposition.

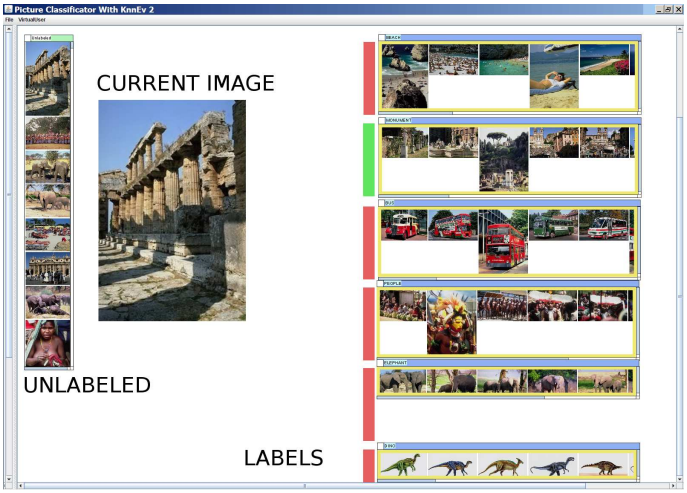

Fig. 3. The GUI. Unlabeled images are in the vertical list and the labeled classes are the horizontal lists. The unlabeled list is sorted by the current strategy selected (in other panel not represented here), and the first one is the current image in the center. After a scrutinising time, the image moves to the most positive class.

5.2. Typical User scenario

During a session, the user may choose the different strategy. A typical User Scenario may be the following:

1. **zero page** one image x_0 is chosen randomly in X.

Database	Label	Number	Comment
gt61	Elephant	50	heterogeneous
	Dinosaur	100	homogeneous
	Bus	100	homogeneous
	Beach	100	composite
	Place	100	very heterogeneous
	People	46	composite
gt62	Elephant	50	heterogeneous
	Flower	100	homogeneous
	Food	100	composite
	Horse	100	homogeneous
	Mountain	100	homogeneous
	People	46	composite

Table 1. The ground truth of two subsets selected from the Corel database. "Homogeneous" means that the visual content is very similar. "Heterogeneous" means that images can have very different content and they are often linked only by the semantic. "Composite" means a mix of the two cases.

2. MR strategy: the most dissimilar images are presented to the user and he decides to define a new class or to put it in an existing class according to the multi-class constraint (section 2).
3. MP strategy: user wants to grow up the classes for reinforcing the visual content description.
4. MC strategies: adding a lot of images may increase the ambiguity between classes because the classes can be potentially superposed in the feature space. The MC strategies will propose the more difficult at class boundary.

5.3. Experimental protocol

5.3.1. Database and visual features

The experiments below are executed on two subsets from the well-known Corel database described table 1. Each set contains 6 classes. Some of these classes, such as "dinosaur", "flower", "bus" may have very similar visual content. Others classes like "mountain", "beach", "place", "elephant" and "people" have more various ones. The distance used is a L2 between a vector containing a standard 64-bin histogram description in the lab color space and a histogram of 8 orientations with 8 intensities [15].

The goal of the experiment is not to study the behavior of a particular dissimilarity measure, or the quality of the visual feature description, and not only to observe which strategy gives the lowest error rate, but to analyse at which steps the errors are produced.

5.3.2. User simulation

First, a ground truth has been built with our supervised system for each of the two sets of data. Two images by classes have been selected randomly to initialize the zero page. Then, the system replays the classification process automatically. At each step, a image is selected by the current strategy, and if the suggestion of the classifier doesn't match the ground truth, an error is detected and counted, and the image is pushed in the correct class. With these experiments, we simulate a virtual user who makes the corrections that a real user would do. Graphics are produced in order to analyze the effects of the strategy on the final lowest error rate, as well as the progression of the error count during the whole process.

5.4. Result and analysis

The graphic 4 shows the evolution of the error number at each step of the classification on the first dataset. Six different strategies are compared: MP, LP, MR, and the local ambiguous strategies for 2 classes MLC2 and LLC2. The final error rate score is between 21.1% and

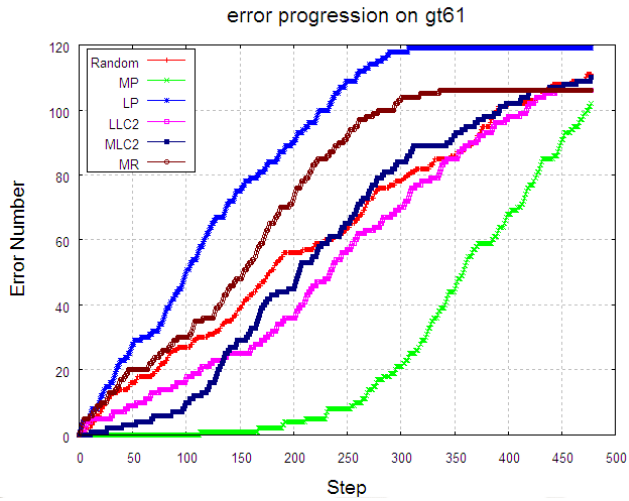

Fig. 4. Error number by step of classification from 6 different strategies on *gt61*. One can notice 3 kind of distinctive behavior: reinforcement for MP, generalization for LP and MR, and mostly linear for the local ambiguous strategies.

24.6%. It means that, for this database, the classifier gives, in best case, about one wrong suggestion for 3-4 good decisions. If we compare the "best" and the "worse" strategy, we observe a difference of 20 errors, which may be significative for 496 images classified.

However the most interesting part is the behavior of the strategies during the whole process. The graphic 4 shows that there are 3 distinctive behaviors. A reinforcement behavior is observed for MP. This strategy begins by labeling one third of the set without errors. Then, the error rate increases progressively to reach at the end the lowest error rate of all the available strategies. We may say that this strategy is robust and minimize the global error rate by minimizing the risk of the decision.

A generalization behavior concerns the LP and MR. These two strategies select different kind of images, the ones with the most diversity of content for MP, and the ones with the lower decision weight according to the classifier. But the two strategies give a generalizing process. It means that they ask a lot of effort to the user at the beginning, but in the last third part, only few work is asked to the user and the number of errors keeps unchanged or constant.

A mostly linear behavior is observed for the ambiguous strategies. For the MLC strategy we may observe that the linear behavior tends to look like a "S". Errors occur sometimes at the beginning, more often in the second third and less at the end of the process.

These three behaviors are observed also for the second set of data *gt62*. But, on this second database, the behaviors are amplified. There is a significant difference of 25 errors at the end between the "best" and the "worse" strategies. The MR strategy seems to be the best compromise between generalization and final error number. The "S" behavior of the MLC is accentuated. Finally, considering the experiment on the ground truths, to optimize the global error rate, the MP strategy can be chosen. A user be on hurry and brave will

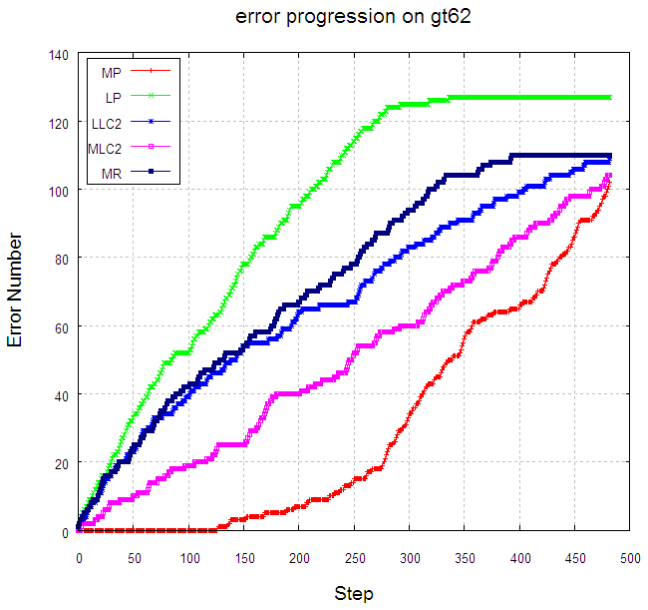

Fig. 5. Error number by step of classification from 6 different strategies on *gt62*. This experiment confirm the previous behavior for on *gt62*.

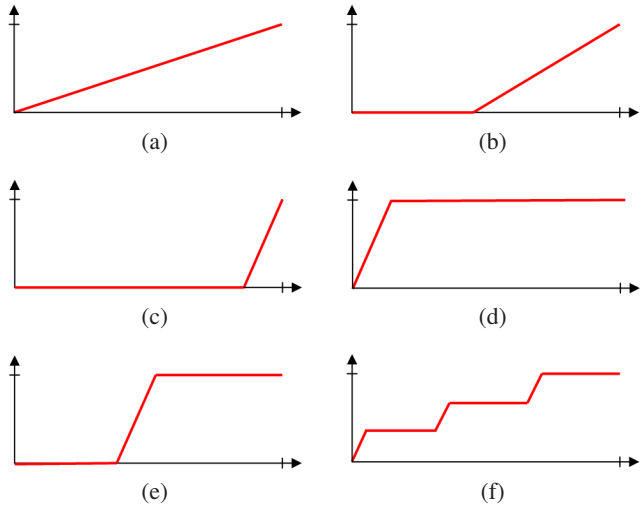

Fig. 6. Typical error profiles.

take a MR strategy and the system may alert him when no error is made from a long time to switch in a full automatic process.

These graphics may be correlated with the cognitive charge of the user. Let considers a work session where a user must classify a collection about several hundred images, which it seems to be a realistic situation at INA. Nowadays it's seems impossible to propose a perfect classifier which make no errors because of the semantic gap. Indeed, the main question is how much and when asking effort to the user. Figure 6 takes an inventory of the characteristic profiles of the errors. The first one (a), the linear one doesn't seem to be very interesting because the classifier makes wrong suggestions all the time and the user must pay attention during the whole session. The two next (b) and (c) characterise the reinforcement strategy which involve the user attention at the end of the session. Selecting the Most Positive at the beginning is useful to help the user to perceive the content of the classes. A generalizing strategy (d) like MR or LP could be efficient ones but, as seen in the previous experiment, they generate a higher error rate then more effort is asked to users. A perfect "S" strategy (e) would be a good compromise by taking advantages from the reinforcement and generalizing strategies.

Finally, if there are a lot of images to classify, a composite strategy (f) could be a good solution to alternate peak of user attention and inaction phases. This hypothesis will be analysed by considering cognitive process in future works.

6. CONCLUSION AND FUTURE WORKS

In this paper, we have built a formal framework of decision. We have defined jointly an evidential classifier and some sampling strategies using the Transferable Belief Model. Thanks to this framework we are able to express intuitively different sampling strategies and to test them easily. We demonstrate that we have three kinds of behavior which generate errors at different steps of the classification process. The Less Positive strategy appears to have interesting capacity for a generic process, such as the Most Reject strategy. The Most Positive strategy minimizes the final error rate and may gain the confidence of a user by doing only few errors at the beginning of the process.

In future work, we will define different BBA, notably by modeling the doubt.

We may also improve the classifier by refining parameter estimation like σ, α_0 and k, or by testing another kernel function and descriptor distances.

Moreover, we are confident in pushing down the error rate by combining strategies. Considering the Corel database, we find in our experiments that only 7% of the images generates errors for all strategies. Indeed, theoretically it should be possible to decrease considerably the error rate, maybe by alternating or combining strategies like in [16], or by defining other strategies.

7. REFERENCES

[1] Bertrand Le Saux and Nozha Boujemaa, "Unsupervised categorization for image database overview," in *Visual Information and Information Systems*, 2002, pp. 163–174.

[2] N. Grira, M. Crucianu, and N. Boujemaa, "Unsupervised and semi-supervised clustering: a brief survey," Report of the MUSCLE European Network of Excellence (FP6), 2004.

[3] N. Grira, M. Crucianu, and N. Boujemaa, "Fuzzy clustering with pairwise constraints for knowledge-driven image categorization," in *IEEE Proceedings on Vision, Image and Signal Processing*, 2006.

[4] Deok-Hwan Kim, Jae-Won Song, Ju-Hong Lee, and Bum-Ghi Choi, "Support vector machine learning for region-based image retrieval with relevance feedback," *ETRI Journal*, vol. 29, pp. 700 – 7002, Oct 2007.

[5] R. Veltkamp and M. Tanase, "Content-based image retrieval systems: A survey," 2000.

[6] N. Vasconcelos and M. Kunt, "Content-based imageg retrieval from image databases: Current solutions and future directions," in *ICIP01*, 2001, pp. III: 6–9.

[7] Edward Chang, Simon Tong, Kingsby Goh, and Chang-Wei Chang, "Support vector machine concept-dependent active learning for image retrieval," in *IEEE Transactions on Multimedia*, 2005.

[8] Michel Crucianu, Marin Ferecatu, and Nozha Boujemaa, "Relevance feedback for image retrieval: a short survey," Report of the DELOS2 European Network of Excellence (FP6), 2004.

[9] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *ninth ACM international conference on Multimedia*, 2001.

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

[11] Thierry Denoeux, "A k-nearest neighbor classification rule based on dempster-shafer theory," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 25, pp. 804–813, 1995.

[12] Glenn Shafer, *A Mathematical Theory Of Evidence*, Princeton University Press, 1976.

[13] Smets Ph., "Belief functions : the disjunctive rule of combination and the generalized bayesian theorem," *J. Approximate Reasoning*, pp. 1–35, 1993.

[14] Smets Ph., "Decision making in the tbm: the necessity of the pignistic transformation," *J. Approximate Reasoning*, pp. 133–147, 2005.

[15] Marin Ferecatu, *Image retrieval with active relevance feedback using both visual and keyword-based descriptors*, Ph.D. thesis, University of Versailles Saint-Quentin-en-Yvelines, 2005.

[16] Yi Wu, I. Kozintsev, J.-Y. Bouguet, and C. Dulong, "Sampling strategies for active learning in personal photo retrieval," in *IEEE International Conference on Multimedia and Expo*, 2006, pp. 529–532.