Multi-labeled image classification by TBM active learning

Hervé Goëau
Institut National l'Audiovisuel
Bry-sur-Marne
France
Email: herve.goeau@inria.fr

Michèle Rombaut, Denis Pellerin
GIPSA-Lab
Grenoble University
France
Email: firstname.author@gipsa-lab.inpg.fr

Olivier Buisson
Institut National de l'Audiovisuel
Bry-sur-Marne
France
Email: obuisson@ina.fr

***Abstract*—The Transferable Belief Model (TBM) is well adapted for knowledge representation, especially for complex systems. In this contribution, the TBM is used as a basic tool of an assistance system for image collection classification. The first part of the system, which is completely automatic, models all available knowledge provided by the already labeled images in order to structure the unlabeled ones. The second part is a user assistance system that proposes an ordered list of images to be labeled according to a specific strategy as well as a possible label. Via a suitable interface, the user agrees or not with the proposal and the global knowledge is updated.**

Keywords: Classification, active learning, image processing

I. Introduction

The transferable belief model (TBM) is particularly well suited to the modeling of knowledge (or lack of knowledge) about a complex system. In the framework of classification, knowledge modeling is a crucial point. Many applications are based on probabilistic tools, but more recently new algorithms have been developed with a TBM approach [1]–[3]. In fact, the probabilistic approach for classification is very efficient when it is possible to use a large enough learning database which is representative of the classes. However, the establishment of such a learning database is a laborious task for the user. An other problem is the semantic interpretation of the classes. Sometimes, the user gives the same interpretation (same class) to two objects with largely different characteristics. However, some very similar objects can be differently interpreted due to a small detail of the objects. This means that for complex classification tasks, especially when extracted characteristics are not enough adapted to differentiate two close classes, it is necessary to use the users'skills. The goal of the classification system is then to exploit the knowledge of the user, but to limit his task as much as possible.

This paper focuses on the problem of image classification. The initial database belongs to INA[1] corporation which preserves broadcast videos but also images concerning scene filming. In order to enlarge the experimentation, other databases were used such Corel database or personal image databases (wedding, holidays, ...). The system we have developed is an assistance classification system. It is based on the fact that it is difficult for a user to a priori define all the classes and manage all the images of the database simultaneously. The main idea is then to select images for the user which are "interesting" to classify according to a specific strategy and possibly to propose a label. The user can accept the proposed label and do nothing, or change or create a class. The automatic image selection is carried out from the accumulated knowledge from the previous image classification.

[1]INA: French Audiovisual Institute

II. Architecture of the proposed system

The framework is divided into two main parts [4]: a fully automatic part for "modeling the knowledge", and another part which concerns the user interactions which select the images to be labeled via a graphic and convivial user interface. It is presented as three modules (Figure 1).

The first module ("multi-labeled classification") exploits the knowledge on the classes given by the already labeled images to characterize the knowledge on the unlabeled images. For each image, characteristics are extracted and compared to those of the labeled images. So, information concerning which class the images belong to is modeled on belief functions.

Then, in the interactive part, the user has to classify the unlabeled images. The second module ("active sampling") selects unlabeled images according to a particular strategy in order to focus the user's attention on them with priority given to one or a small set of images. The user can choose and switch to the most relevant strategy. For instance, he might be interested as a priority in labeling images with visual contents which are relatively close to one class of images. Later, he may be more interested in the labeling of "difficult" images, for instance with such different visual content from the known classes that visual diversity has to be explored. On a graphic user interface (GUI) (third module), the label proposal are automatically carried out on the images selected. Then, the user has to accept, correct or complete (in multi-label cases) the proposal. When the selected images are labeled, the knowledge modeling of the first module is updated in order to select more significant unlabeled images and to improve the accuracy of the label proposal in the following rounds.

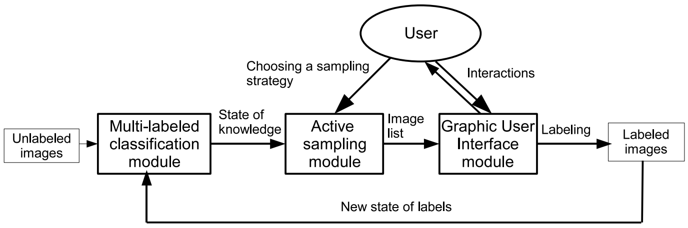

Figure 1. Representation of the global system

III. Multi-labeled classification module

In this part, the system gives a multi-label classification for each unlabeled image, according to the knowledge given by the already labeled images. The pre-processing step consists of characteristic extraction for each image (orientation and color histograms). This step can be done only once before the main processing step.

The first main part consists in modeling the knowledge of the labeled images in order to predict the relevant label of the current unlabeled image. This step is a combination of different knowledge extracted from (i) the neighbor images of the current image, (ii) the known classes and (iii) image characteristics. First, the knowledge is modeled from a neighboring image of the current image, then from the k nearest neighbors belonging to the same class, then from all the classes, then from all the characteristics.

A. Knowledge from the neighbor image

Considering one class C_q, leading to the frame of discernment Ω_q:

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

where hypothesis H_q means "the image u belongs to the class C_q" and $\overline{H_q}$ means the opposite hypothesis. Assuming that two visually similar images generally tend to have the same label, if the unlabeled image u is close to the labeled image l_q^i (small distance $d(u, l_q^i)$ in the feature space), a high belief is assigned to the proposition H_q while a low belief is associated to the doubt (with a lower bound of $1 - \alpha_0$). We choose to use the Basic Belief Assignment (BBA) $m_i^{\Omega_q}$ proposed in [5]:

$$\begin{aligned} m_i^{\Omega_q}(H_q) &= \alpha_0 . e^{-\left(\frac{d(u,l_q^i)}{\sigma_q^i}\right)^\beta} \\ m_i^{\Omega_q}(\Omega_q) &= 1 - m_i^{\Omega_q}(H_q) \end{aligned} \tag{2}$$

This model is very interesting when a class is represented by several modalities in the characteristics space. This means that two distant images in this space can still belong to the same class. It can be noticed that for a particular class C_q, the BBA form proposed will not generate conflict.

B. Knowledge from K nearest neighbors

This belief in the proposition H_q can possibly be strengthened by the other image samples of a same class C_q. Following the method described in [5], the BBA from the K nearest neighbors images $\{l_q^0, l_q^1, \ldots, l_q^k\}$ of the class C_q are conjunctively combined in order to update the belief functions:

$$\begin{aligned} m^{\Omega_q}(\Omega_q) &= \prod_{i=0}^{k-1} (1 - \alpha_0 . e^{-\left(\frac{d(u,l_q^i)}{\sigma_q^i}\right)^\beta}) \\ &= \prod_{i=0}^{k-1} (1 - m_i^{\Omega_q}(H_q)) \\ m^{\Omega_q}(H_q) &= 1 - m^{\Omega_q}(\Omega_q) \end{aligned} \tag{3}$$

Once one nearest neighbor gives a high value to H_q hypothesis, then the belief that image u belongs to this class is high. Otherwise, if all the neighbors are far from u, then belief functions give a high value to doubt, and also to the combined BBA.

C. Mass transfer

Up to this step, no mass value is associated with the proposal $\overline{H_q}$ and there is no conflict between BBAs. Indeed, it is advisable to discern positive from negative labels in order to help the user to choose the most relevant labels for one unlabeled image u. This a priori knowledge can be taken into account by an operation which involves transferring the mass values of the propositions H_q and the doubt $\{H_q, \overline{H_q}\}$ to the three propositions H_q, $\overline{H_q}$ and Ω_q. That means if all the K nearest neighbors are far from image u in the characteristics space (large mass on Ω_q), this image probably does not belong to class C_q. It is interesting to transfer the mass at this step in order to avoid conflict that could appear in previous steps. Further, this transfer has a real semantic meaning. The proposed mass transfer is built using a set of three triangular functions controlled by one parameter m_0(see figure 2) to compute the new set of belief functions $m_s^{\Omega_q}$ and verifying the conditions:

$$\begin{aligned} & m^{\Omega_q}(\emptyset) && = 0 \\ & m^{\Omega_q}(H_q) + m^{\Omega_q}(H_q, \overline{H_q}) && = 1 \quad \text{if} \quad m^{\Omega_q}(H_q) > m_0 \\ & m^{\Omega_q}(\overline{H_q}) + m^{\Omega_q}(H_q, \overline{H_q}) && = 1 \quad \text{if} \quad m^{\Omega_q}(H_q) \leq m_0 \end{aligned} \tag{4}$$

This last operation gives a mass distribution $m_s^{\Omega_q}$ quantifying the beliefs for one unlabeled sample u related to one class C_q. The parameter m_0 is set to 0.5 because, without more a priori information, it corresponds to equal distribution between the two initial proposals.

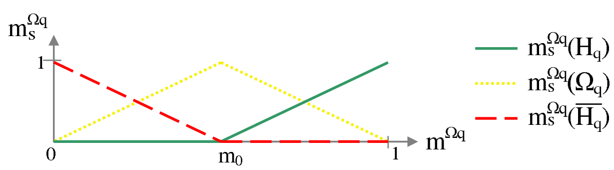

Figure 2. The triangular functions for the transfer mass operation

D. Knowledge from all classes

Depending on the semantic interpretation of the images, classes can be non-exclusive. That means one particular image can simultaneously belong to two (or more) different classes (for instance an image with someone in front of the sea may be put in the two classes associated with the labels "sea" and "people"). It could be interesting to allow the user the opportunity of carrying out a multi-affectation of an image. So the problem is now to compute a new mass distribution quantifying membership of the unlabeled sample u to none or several classes. Considering a set of Q classes $C = \{C_1, C_2, \dots C_Q\}$, a new frame of discernment Ω is defined as the product space of the local frames of discernment Ω_q previously defined for each class C_q:

$$\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_Q \quad (5)$$

If some classes of C are exclusive, this a priori knowledge can be used to simplify the set Ω. Considering one unlabeled image u, each class C_q provided a belief function $m_s^{\Omega_q}$ in its own frame of discernment Ω_q. Given a proposition $(B_1, B_2, \cdots, B_Q)$ of Ω, where a symbol B_q represents one of the three propositions from the local frame of discernment Ω_q, the empty set extension operator [6] computes the new mass m^{Ω} using the following expression:

$$m^{\Omega}(B_1, B_2, \cdots, B_Q) = \prod_{B_q \in 2^{\Omega_q}}^{Q} m^{\Omega_q}(B_q) \quad q \in [1, Q] \quad (6)$$

One can note the cost involved by this operator. The global cost of computing is reduced in practice by the fact that a local mass distribution m^{Ω_q} only contains two focal elements.

E. Knowledge from all characteristics

Numerous methods of classification use an early fusion of features by concatenating the descriptors extracted from one image in one single vector. One advantage of this approach is the use of a simple fusion model involving a low cost computing compared to a late fusion model. One disadvantage is that it is difficult to express and quantify the discordance between the information given by the features. Besides, the concatenation does not take into account the size of feature vectors. It is extremely difficult to add a new characteristic to the system without a re-programmable phase.

So, we propose to use a late fusion approach in order to solve these problems. Each feature space d_i brings a mass distribution $m_{d_i}^{\Omega}$ describing the beliefs about the membership of one unlabeled sample u to the classes. Then, the conjunctive operator is used to combine all the belief functions computed in the feature spaces considered. The latter operation can introduce part of the mass distribution on the conflict $\emptyset$. This can be useful to detect a new unknown class of images or a new modality of a known class for instance. A caution rule of combination could be used in order to take into account the dependence of the characteristics. However, the final mass is only used in the sampling process as seen in section IV-A to compare the different unlabeled images.

At the output of the knowledge modeling module, an unlabeled image u is associated to a mass distribution m^{Ω} quantifying the belief about the known class taking into account all the extracted characteristics.

F. Interpretation of hypothesis

Because the different classes can be non-exclusive, it is important to interpret the hypotheses belonging to $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_Q$

Positive hypothesis: A positive hypothesis ω_p^q is composed of only one local positive hypothesis such as H_q, the others corresponding to local negative ones such as $\overline{H_n}$:

$$\omega_p^q = (H_q, \overline{H}_{n_1}, \overline{H}_{n_2}, \cdots, \overline{H}_{n_N}) \quad (7)$$

This positive hypothesis ω_p^q can mean that the unlabeled image u belongs to the single class C_q.

Reject hypothesis: This hypothesis ω_r corresponds to all negative local hypotheses such as $\overline{H_q}$:

$$\omega_r = (\overline{H_1}, \overline{H_2}, \cdots, \overline{H_Q}) \quad (8)$$

This means that all Q known classes do not correspond to the unlabeled image u. It can be used to initiate a new class, or new visual content of a known class.

Ambiguous hypothesis: The set of other hypotheses corresponds to a multi-class category. These hypotheses are composed of P (with $P \geq 2$) local positive hypotheses such as H_q and N (with $N \leq Q-2$ and $P+N=Q$) local negative hypotheses such as $\overline{H_q}$. The general form of such a hypothesis ω_a^P is:

$$\omega_a^P = (H_{p_1}, H_{p_2}, \cdots, H_{p_P}, \overline{H_{n_1}}, \overline{H_{n_2}}, \cdots, \overline{H_{n_N}}) \quad (9)$$

The degree of ambiguity P of hypothesis ω_a^P corresponds to the number of local positive hypotheses. This means the unlabeled image u can belong to P classes $\{C_{p_1}, C_{p_2}, \cdots, C_{p_P}\}$ simultaneously, but not to the other classes $\{\overline{C_{n_1}}, \overline{C_{n_2}}, \cdots, \overline{C_{n_N}}\}$.

In this set of ambiguous hypotheses, the global ambiguous hypothesis ω_{ga} is defined by:

$$\omega_{ga} = (H_1, H_2, \cdots, H_Q) \quad (10)$$

This hypothesis means that the unlabeled image u can potentially belong to all classes. It is not very realistic but it can serve the purpose of concluding that the characteristics are not sufficient.

IV. Active sampling module

It could be difficult for a user to classify a set of images, particularly when this set is large, and when the classes are not defined a priori. This is the case, for instance, when somebody wants to store his holiday images, not only by time stamp, but also by themes (visits, swims, meals, ...). Rather than submitting all the images simultaneously, or one by one in random order, the idea is to propose an "adequate" order following a sampling strategy. So, in this part, a small set of chosen images is proposed to the user to be classified, maybe because they are very similar to labeled images or on the contrary because they are different to labeled images. Active learning is rarely used for multi-labeling [7].

A. Sampling strategies

For all unlabeled images u, a BBA m_u^Ω was computed as was presented in the previous section. It is used to determine the best unlabeled image(s) to be proposed for a labeling step according to a particular strategy.

Most positive unlabeled images (MP): The strategy, some times named "most relevant" [8]), selects the unlabeled image u_{mp} which obtains the highest pignistic probability PP computed on Ω_P [9], subset of Ω made up of only positive hypotheses ω_p^q (eq. 7). It corresponds to the selection of "easy to classify" images, because the visual content is very similar to already labeled images.

$$\begin{aligned} PP(u) &= \max_{\omega_p^q \in \Omega_P} BetP\{m_u^\Omega\}(\omega_p^q) \\ u_{mp} &= \text{argmax}_{u \in U} \ PP(u) \end{aligned} \tag{11}$$

This strategy selects the nearest unlabeled images to the labeled images of the different classes and then improves the knowledge of the classes. However, it does not cover the diversity of the visual content of the image collection.

Most ambiguous unlabeled images: The ambiguity can be global (Most Global Ambiguity MGA) in all the classes of C. This strategy selects the unlabeled image u_{mga} with a maximum of pignistic probability on the proposition ω_{ga} (eq.10) :

$$\begin{aligned} PP(u) &= BetP\{m_u^\Omega\}(\omega_{ga}) \\ u_{mga} &= \text{argmax}_{u \in U} \ PP(u) \end{aligned} \tag{12}$$

This strategy consists in choosing the unlabeled image which is on the borders of all the known classes. The user can affect this image to one or several classes, or create a new class.

It can be interesting to select images that are locally most ambiguous (Most Local Ambiguous MLA_P), meaning they are on the borders of a P class.
Firstly, for each unlabeled image u, the highest pignistic probability is computed on the subset Ω_{LA_P} only made up of propositions ω_a^P (eq. 9) corresponding to local ambiguity between P hypotheses.

$$PP(u) = \max_{\omega_a^P \in \Omega_{LA_P}} BetP\{m_u^\Omega\}(\omega_a^P) \tag{13}$$

Secondly, the most local ambiguous image u_{la_P} is selected by comparing the pignistic probabilities $PP(u)$ of all unlabeled images with P class local ambiguity.

$$u_{la_P} = \underset{u \in U}{\text{argmax}} \ PP(u) \tag{14}$$

For instance, if $C = \{C_1, C_2, C_3\}$, it could be interesting to consider the local borders between two classes. Each unlabeled image $u \in U$ is associated to $PP(u)$ corresponding to the highest pignistic probability on the subset :

$$\Omega_{LA_2} = \{(H_1, H_2, \overline{H_3}), (H_1, \overline{H_2}, H_3), (\overline{H_1}, H_2, H_3)\} \tag{15}$$

Most rejected unlabeled image (MR): In order to explore the visual diversity of the space, it could be interesting to select unlabeled images that do not correspond to any class [10]. The unlabeled image u_{mr} has the highest pignistic probability on the rejected hypothesis ω_r (eq. 8) of the set of discernment Ω:

$$\begin{aligned} PP(u) &= BetP\{m_u^\Omega\}(\omega_r) \\ u_{mr} &= \text{argmax}_{u \in U} \ PP(u) \end{aligned} \tag{16}$$

The user has to decide the correct labels of this image, or has to create a new class with this single new labeled image.

Most Conflicted unlabeled image (MC): The information fusion with all characteristics (color, orientation,...) can lead to a conflict about the inclusion in one or more classes as were presented in section III-E. The unlabeled images with such a high conflict can be interesting to classify, because they do not correspond to current known classes. This conflicting information is directly computed by m_u^Ω of image u.

The most conflicting unlabeled image u_{mc} has the highest $m_u^\Omega(\emptyset)$:

$$u_{mc} = \underset{u \in U}{\text{argmax}} \ m_u^\Omega(\emptyset) \tag{17}$$

Most Uncertain unlabeled image (MU): The information fusion attributes a large mass to the hypothesis of Ω made up of the Ω_q. In terms of pignistic probabilities, the two hypotheses H_q and $\overline{H_q}$ are similar and the BBA is said to be "non-specific". That means it is impossible to distinguish one hypothesis from the others. The *non specificity* $N(m_u^\Omega)$ is computed and the most uncertain image u_{mu} is determined:

$$\begin{aligned} N(m_u^\Omega) &= \textstyle\sum_{\emptyset \neq B \subseteq \Omega} m(B) log_2(||B|) \\ u_m &= \text{argmax}_{u \in U} \ N(m_u^\Omega) \end{aligned} \tag{18}$$

V. Graphic user interface module

A fully automatic classification system is not realistic for a such complex application because the user has an high level semantic interpretation of the images. In order to make the user's task easier, the system can use the knowledge modeled by the belief functions to propose a label for the unlabeled selected image. This can be done, image by image, or by a small set of unlabeled images with the same class.

The interface (shown in figure 3) is an interactive view with images and classes which allows the user to manipulate images. The vertical list of unlabeled images is sorted from the most (top) to the least (bottom) representative image according the current sampling strategy. Only the first images of the list are displayed because of the limited space on the screen and the ability of the user to simultaneously manage several images. Each horizontal list of labeled images represents a class. For the same reason of readability, only a few images are represented.

The first unlabeled image u (the most representative of the current sampling strategy) is duplicated and enlarged between the vertical and horizontal lists. The views of the horizontal lists are updated at each selection of an unlabeled image u by displaying first its k-nearest neighbors from left to right. The proposed classification is then displayed to the user by an arrow pointing to the suggested class. In the case of a multiple-label proposal, multiple arrows are displayed on the suggested classes.

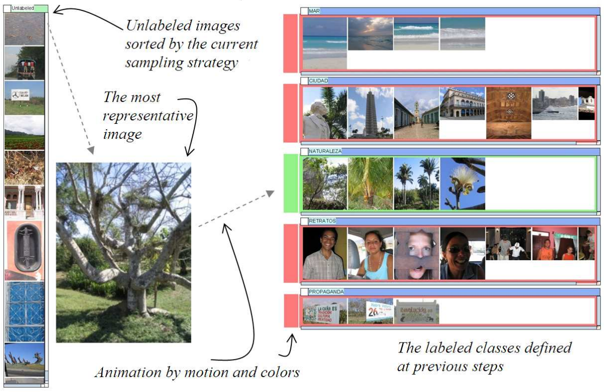

Figure 3. The GUI: unlabeled images are in the vertical list and the classes are the horizontal lists. The unlabeled list is sorted by a given strategy. The most representative image by this sampling strategy is in the center. After a short time, the image moves slowly to the suggested class.

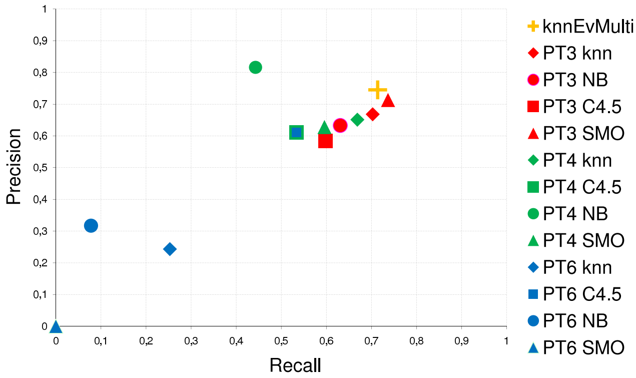

Figure 4. Comparison of different methods of multi-label classification assessed in [13]. The proposed method denoted "KnnEvMulti" is represented by the orange cross.

VI. Experimentations

The evaluation of the system concerns the automatic single and multi labeling and the performance of the different strategies of the active learning process.

A. Automatic classification and multi-labeling

The first part of the system models the knowledge and proposes automatic single or multi-labeling. In order to compare the performances of this system, we use the dataset ("scene-classification") proposed by Boutell in [11] and available on the official website of LibSVM [12]. The images are taken from the Corel image database where the author has identified 6 labels corresponding to concepts ("urban", "sunset", "fall foliage", "field", "mountain" et "beach"). This benchmark contains 2407 images associated each one with between 1 and 5 labels simultaneously (the average is 1.08 labels by image). Each image is described by a vector, the concatenation of the mean and variance of local color histograms. The test consists in training the system with 1211 sample images, and predicting the labels of the 1196 remaining images.

The frame of discernment Ω contains 2^6 hypotheses, all combinations of base hypotheses H_q and $(\overline{H}_q)$ associated to the class C_q $(q \in \{1, 2, \ldots, 6\})$.

The recall and precision measures in the specific case of multi-labeling are used to evaluate classification performance. For each image u_i, Y_i is the number of correct labels from the ground truth and Z_i the number of labels predicted.

The recall index is the ratio of the number of labels correctly predicted to the number of correct labels.

$$Recall(D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \cap Z_i|}{|Y_i|} \quad (19)$$

The precision index is the ratio of the number of labels correctly predicted to the number of predicted labels.

$$Precision(D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \cap Z_i|}{|Z_i|} \quad (20)$$

Experimentally, the best compromise between recall and precision is obtained for a number $k = 5$ nearest neighbors, and a value $f = 0.78$ with L_1 distance: Recall $= 0.714$ and Precision $= 0.745$.

Figure 4 compares our results with 12 classification methods on the same data set presented in [13]. In this paper, the authors compared 4 classification algorithms (kNN [14], C4.5 [15], naive Bayes [16] and SMO (SVM-based [17]) as 3 Problem of Transformation methods named PT3[2], PT4[3] and PT6[4]). The results show the proposed method as being one of the most successful with the "PT3+SMO" method (Recall $= 0.737$ and Precision $= 0.713$). This result validates the relevance of our approach to a difficult problem of classification.

B. Strategy characterization of active image sampling

The system has been designed to help a user to organize a collection of images from the beginning, ie free of any label. There is therefore no basis for learning. The sampling strategies aim to select images to be labeled by the user and complete the basic learning on the fly.

The strategy characterization studies the impact of active sampling strategies on the classification performance. At the beginning, only a few images are labeled. During the experiment, the numbers of incorrect labels according to the ground truth is integrated corresponding to the task of the user. Figure 5 shows the evolution of this parameter during the whole experiment. The horizontal axis represents the sequence of successive selections of images, and the vertical axis corresponds to the accumulation of incorrect labels proposed automatically by the system. We use a Corel image database of 500 images called *gt500*. The database contains 5 classes, each containing 100 images related to different categories: "flowers","bus","dinosaurs","beaches" and "horses". The 5 classes have already been identified by a single sample image. Each image only belongs to one class. 495 images remain to be classified one by one.

[2]Considering each different set of labels that exist in the multi-label data set as a single label

[3]Learning binary classifiers, one for each different label

[4]Using a set of binary classifiers

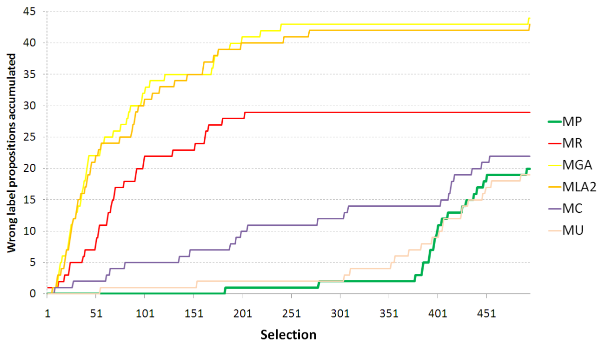

Figure 5. Comparison of the six strategies on the dataset *gt500* under the same experimental conditions (distance d_{Bhat}, descriptors orientation and color in space R, G, B, $f = 0.4$ and $k = 5$). The x axis corresponds to the successive selections of images, the y axis corresponds to the accumulation of wrong automatic classifications.

Six active sampling strategies have been tested: most positive (MP), most rejected (MR), most global ambiguity (MGA), most local ambiguity in 2 classes (MLA2), most conflicted (MC) and most uncertain (MU) as presented section IV-A. The strategies are tested individually. An image is selected according to the strategy chosen, then a labeling proposal is performed on this image. If this proposal does not match the ground truth simulating a user's selection, it counts the number of incorrect labels accumulated, then adds that image to the training set with the right label.

The curves can be read in 2 ways: (i) The last point of the curve allows the final performance of classification involved the use of a strategy to be focused on. (ii) By observing the curves entirely, it is possible to see when wrong proposal labeling occurs.

In this experiment, the strategies can be separated into 2 groups: first group for which wrong proposal labeling is delayed (most positive (MP), most conflicted (MC) and most uncertain (MU)), second group for which the wrong proposal labeling occurs early in the classification of images.

VII. Conclusions

The belief functions are particularly well suited to modeling the knowledge about image classes when the learning base is poor, especially in the beginning of the classification process. At anytime, the system proposes one or more possible label for each unlabeled image, by comparing this unlabeled image with its K nearest neighbors of different known classes. The belief model is well adapted to the description of the different strategies of unlabeled image presentation. The last figure clearly shows the characteristics of these strategies which ask the user to carry out important task in the beginning or at the end of the classification process. Now it would be interesting to study a possible combination of strategies over time in order to limit the task of the user and to improve the final classification performance.

We thank the Rhône-Alpes region for their support with LIMA project, and the INA Institute for supporting the grant of H. Goëau.

References

[1] T. Denoeux and P. Smets, "Classification using belief functions: the relationship between the case-based and model-based approaches," *IEEE Transactions on Systems, Man and Cybernetics B*, vol. 36, no. Issue 6, pp. 1395–1406, 2006.

[2] P. Vannoorenberghe, P. Smets, and L. Godo, "Partially supervised learning by a credal em approach," in *ECSQARU 2005. LNCS (LNAI)*, vol. 3571. Heidelberg: Springer-Verlag, 2005, pp. 956–967.

[3] Z. Younes, F. Abdallah, and T. Denoeux, "An evidence-theoretic k-nearest neighbor rule for multi-label classification," in *Proceedings of the 3rd International Conference on Scalable Uncertainty Management (SUM 2009)*. Washington, DC, USA: Springer-Verlag, September 2009, pp. 297–308.

[4] H. Goëau, "Structuration de collections d'images par apprentissage actif crédibiliste," PhD, Université Joseph Fourier, Grenoble, France, May 2009.

[5] T. Denoeux, "A k-nearest neighbor classification rule based on dempster-shafer theory." in *IEEE Trans. on SMC*, 1995, pp. 804–813.

[6] P. Smets, "Belief function: the disjonctive rules of combination and the general bayesian theorem," *International Journal of Approximate Reasoning*, vol. 9, pp. 1–35, 1993.

[7] G.-J. Qi, X.-S. Hua, Y. Rui, J. Tang, and H.-J. Zhang, "Two-dimensional multi-label active learning with an efficient online adaptation model for image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 99, no. 1, 2008.

[8] M. Crucianu, M. Ferecatu, and N. Boujemaa, "Relevance feedback for image retrieval: a short survey," Report of the DELOS2 European Network of Excellence (FP6), 2004. [Online]. Available: citeseer.ist.psu.edu/crucianu04relevance.html

[9] P. Smets, "Decision making in tbm: the necessity of the pignistic transformation." *Journal of approximate Reasonning*, vol. 38, no. Issue 2,, pp. 133–147, February 2005.

[10] K. Brinker, "Incorporating diversity in active learning with support vector machines," in *Proceedings of the 20th International Conference on Machine Learning*. AAAI Press, 2003, pp. 59–66.

[11] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern recognition*, vol. 37, no. 9, pp. 1757–1771, 2004.

[12] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[13] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *International Journal of Data Warehousing and Mining*, vol. 3(3), pp. 1–13, 2007.

[14] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, no. 1, pp. 37–66, 1991.

[15] J. R. Quinlan, *C4.5: programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[16] G. H. John and P. Langley, "Estimating continuous distributions in baysian classifiers." in *Eleventh Conference on Uncertainty in Artificial Intelligence*, P. Besnard and S. Hanks, Eds., Montreal Canada, 1995, pp. 338–345.

[17] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization." C. B. B. Schölkopf and A. Smola, Eds. MIT Press, 1999, pp. 185–208.