Violent Web images classification based on MPEG7 color descriptors

Radhouane Guermazi*, Mohamed Hammami† and Abdelmajid Ben Hamadou*

Multimedia InfoRmation system and Advanced Computing Laboratory

*Higher Institute of Information Technologie and Multimedia, Sfax-Tunisia

Email: rguermazi@laposte.net, abdelmajid.benhamadou@isimsf.rnu.tn

†Faculty of Sciences, Sfax-Tunisia

Email: mohamed.hammami@fss.rnu.tn

Abstract—In this article, we present a contribution to the violent Web images classification. This subject is deeply important as it has a potential use for many applications such as violent Web sites filtering. We propose to combine the techniques of image analysis and data-mining to relate low level characteristics extracted from the image's colors to a higher characteristic of violence which could be contained in the image. We present a comparative study of different data mining techniques to classify violent Web images. Also, we discuss how the combination learning based methods can improve accuracy rate. Our results show that our approach can detect violent content effectively.

Index Terms—content-based image clasification, Data-Mining, MPEG 7, violence

I. INTRODUCTION

The World Wide Web has had an impact on nearly every aspect of society including commerce, science, politics and government, personal relationships, and health. However, the Web has its downside. When surfing it's easy to come across sites containing litigious content such as violent content. Indeed, kids are exposed to a continuum of violence on the Internet, ranging from sites with sophomoric cruel humor to disturbing depiction of torture and sadism. Today, children and teens can download violent music lyrics (including lyrics removed from retail versions of songs) and access violent images and video clips, with a click of a mouse, Young people use the Web to explore and display their creativity in very exciting ways creating online movies, music, Web sites and blogs or online diaries.

Our initial work [1], has proposed a violent Web content detection and filtering system called "WebAngels filter" which uses textual and structural analysis. Our experimental tool evaluation shows its effectiveness to detect and filter Web violent pages, nevertheless, "WebAngels filter" finds difficulty in classifying the violent sites that contain only graphic violent content. In this paper, we investigate this problem and we propose a novel approach of violent Web images classification. We focus our attention on the use of colors descriptors and classifiers combination and we demonstrate that it can be applied to improve the violent Web images classification accuracy rate.

The remainder of this article is organized as follows. In section II, we started out with over viewing related work according to images classification and retrieval. Following that, the proposed approach of violent Web images classification is reviewed in section III. An intensive experimental evaluation and comparison results are discussed in section IV. In the following section, we show the efficiency of combining data mining techniques to improve the classification accuracy rate. Finally section VI summarizes our contributions and presents some concluding remarks and future work directions.

II. IMAGES CLASSIFICATION

The images classification consists of systematically distribute images according to established classes in advance. Classifying an image, makes a class match to it, marking its relationship with other images. In general, recognizing an image is an easy task for a human being. Indeed, throughout his experience, he gained knowledge which enables it to adapt to changes resulting from different acquisition conditions. It can for example recognize an object in several directions, partly hidden by another, near or far, and on various illuminations. However, technological progress in terms of acquisition image (cameras, sensors, microscopes) and storage cause rich database information and growing areas of applications make manual images annotation with text descriptors impractical for two reasons: the first is that a considerable level of human labor is required for manual annotation. The second is the annotation inaccuracy due to the subjectivity of human perception. Indeed, It is difficult for two persons to describe the same image with the same text [2]. The images annotation must use additional means , other keywords, such as a content image description based on visual descriptors.

The Descriptors (or image signature) are a n-dimensional vectors automatically extracted using computer vision techniques. The best choice of descriptors, which faithfully summarize the visual and semantic image content, is an extremely important issue. Indeed, the quality of the results is highly dependant on the quality of the used descriptors. The most commonly visual features used to describe the image content are color defined on a selected color space serving for different applications. Common color features or descriptors include, color-covariance matrix, color histogram, color moments, color coherence vector, dominant color, color structure, scalable color, and color layout [3]. Texture provides important information in image classification as it describes

the content of many real-world images such as skin, trees, and bricks. Hence, texture is an important feature in defining high-level semantics for image classification. Texture features commonly used in image retrieval systems include spectral features, such as features obtained using Gabor filtering or wavelet transform [4]. Shape is a fairly well- defined concept. Shape features of general applicability include aspect ratio, circularity, Fourier descriptors, moment invariants, consecutive boundary segments [5].

Many researches have emerged to extract and interpret the low level image features. The main difficulty encountered by the developed techniques is the "semantic gap" that represents the discrepancy between the limited descriptive power of low-level image features and the richness of user semantics.

Several techniques have been developed to reduce the "semantic gap" using learning methods that can be classified into two categories: (1) supervised learning methods and (2) unsupervised learning methods. The goal of supervised learning is to predict the value of an outcome measure based on a set of input measure. In unsupervised learning, there is no outcome measure, and the goal is to describe how the input data are organized.

Among the supervised learning methods, we cite the Support Vector Machine (SVM) used by Zhang et al. to categorize images judged relevant and irrelevant by the user in a relevance feedback retrieval system [6]. Vailaya et al. used a binary Bayesian classifier, high-level concepts of natural scenes are captured from low-level image features. Database images are automatically classified into general types as indoor/outdoor, and the outdoor images are further classified into city/landscape... [7]. In [8], the authors use a neural network to establish links between low-level features extracted from images regions and 11 concepts. Besides the algorithms mentioned above, decision tree are also used to derive semantic features. For example, Sethi et al. use the CART algorithm to derive decision rules mapping global color distribution (HSV space color histogram) in a given image to textual description (four keywords: Sunset, Marine, Arid images and Nocturne). [2]. In [9], C4.5 was used to build a decision tree based on a set of images relevant to the query, and then used as a model to classify database images into two classes: relevant and irrelevant. This algorithm is used in the Relevance Feedback loop to provide relevant images for the user to label at next iteration. A similar methodology is employed on [10] where the ID3 decision tree has been used to classify the images as relevant/irrelevant based on their color features,

As supervised learning, there are several approaches based on an unsupervised methods to classify and research images. For example, Chen et al., proposed a method, named CLUE, to retrieve semantically coherent image clusters. Their method is based on the NCut clustering to cluster the target images into different semantic classes [11]. Shi et al. attempt to resolve the inefficiency of the traditional classification methods, like k-means to classify images with different concepts, by proposing a spectral clustering method Normalized cut (NCut) [12].

Aiming to classify Web images to violent/non-violent im-

ages, we have been led to reflect on the approach to achieve this goal. To that finality, we opted for a compared study of the use of several supervised methods powered by a set of MPEG7 colors features. In the following, we present our approach to violent Web images classification.

III. VIOLENT WEB IMAGES CLASSIFICATION APPROACH

We propose to build an automatic violent content detection solution based on a machine learning approach using a set of manually classified Web images, in order to produce a prediction model, which makes it possible to know which images are violent and which are not. Figure 1 illustrates the general schema of proposed approach.
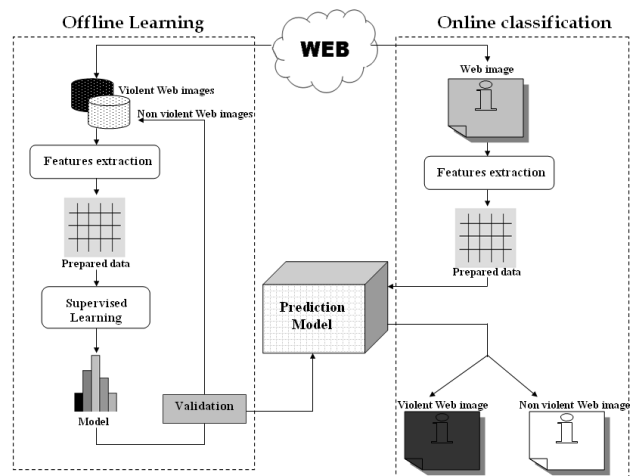

Fig. 1.   General schema of proposed approach.

Two parts can be distinguished: the first one serves to offline learning consacred to the predicted model preparation. The second one serves to online Web images classification. To classify the images into two classes, we are based on the KDD process to extract useful knowledge from volumes data [13]. The general principle of the approach of classification is the following: Let I be the population of images to be classified. To each image i of I one can associate a particular attribute, namely, its class label C. C takes its value in the class of labels (0 for violent, 1 for non violent)

$$C : I \rightarrow \Gamma = \{Violent, nonViolent\}$$
$$i \in I \mapsto C(i) \in \Gamma$$

Our study consists in building a means to predict the attribute class of each Web image. To do it, three major steps are necessary: (a) data preparation by the extraction of visual content based features; (b) supervised learning; (c) evaluation and validation of the learnt model on the training and test data set.

A. Data Preparation

In this stage we identify exploitable information and check their quality and their effectiveness in order to build a two-dimensional table from our training corpus. Each table row

represents a Web image and each column represents a feature. In the last column, we save the Web image class (0 or 1).

1) Learning and test data set: The data mining process for violent Web images classification requires a representative learning data set consisting of a significant set of manually classified Web images. Considering the diversity and the enormous number of Web images on Internet, this phase constitutes a considerable work.

Our corpus is a typical set of 1000 Web French and English pages 500 of which are violent [14]. In this article we are interested in violent image classification, so we first extract images from the Web pages by examining the img tags. Note that the extracted images are manually classified into violent and non violent class after discarding small icons such as buttons and rulers used for alignment, segmentation. Our data base contains a set of 1787 images, 895 of them are violent. For the first category, we find images reflecting war, horror, prisoners torture, murder and explicit violence . For non-violent images, we find confusing images, especially images containing the colors that can exist in the violent images such as red and dark colors. The rest of images contains images of faces, plants, etc. 70% of the images are used for the learning, the rest for the test.

2) Visual content-based features: Color is one of the most important and easily identifiable feature for describing visual content. In our case, the violent Web image classification, the color descriptors seem to be the most relevant to better discriminate the violent Web images from those inoffensive. Indeed, during the data base collection, we noticed that the majority of violent images contains a number of a specific colors like red that marks generally the blood existence. Also violent images are darker.

In the extraction of visual features, we are based on the The MPEG7 color descriptors [15], to know dominant color; layout color and scalable color, that cover different aspects of this important visual feature. In this section we explain the application of these descriptors to construct the vector of visual characteristics used to classify the Web images.

a) Dominant Color Descriptor (DCD): is a set of dominant colors in a region of interest or in an image providing a compact description that is easy to index. In order to compute this descriptor, the colors present in a given image or region are first clustered using a vectorial quantification like LBG algorithm. This results in a small number of colors and the percentages of these colors are calculated. The variances of the colors assigned to a given dominant color are also computed. A spatial coherency value is also computed that differentiates between large color blobs versus colors that are spread all over the image.

The descriptor is thus defined by

$$DCD = ((c_i, p_i, v_i)_{1 \le i \le n}, CS) \tag{1}$$

where n is the number of dominant colors. It can vary from image to image and a maximum of eight dominant colors can be used to represent image. c_i is the ith dominant color; p_i its percentage value; CS is a single number that represents the overall spatial homogeneity of the dominant colors in the image and v_i is the ith color variance defined by:

$$v_i = \frac{1}{Nb_i} \sum_{k=0}^{k=N-1} (m_i - p_{ki})^2 \tag{2}$$

Where Nb_i is the number of pixels of the ith dominant color in the image, m_i (m_{H_i}, m_{S_i}, m_{V_i}) is the ith domiant color value and p_{ki} is the initial value of the pixel (before quantization) that has the same values of H, S and V with the ith dominant color (after quantization).

In our work, we have prepared a range of 128 values where each value corresponds to a possible combination of H, S and V. We insert the percentages of the eight dominant colors in their corresponding positions. The rest of values is equal to 0. In addition to these values, we calculate 3*8=24 values corresponding to the 8 variances $(v_{H_i}, v_{S_i}, v_{V_i})_{1 <= i <= 8}$ of the dominant colors. Finally, we add the value of spatial coherence.

b) Color Layout descriptor (CLD): is a compact and resolution-invariant MPEG-7 visual descriptor defined in the $YCbCr$ color space and designed to capture the spatial distribution of color in an image. The feature extraction process consists of two parts: grid based representative color selection and DCT transformed with quantization. An input picture is divided into 64 blocks and their average colors are derived. Note that it is implicitly recommended that the average color be used as the representative color for each block. This partitioning process is important to guarantee the resolution or scale invariance. The derived average colors are transformed into series of coefficients by performing DCT. A few low-frequency coefficients are selected using zigzag scanning and quantized to form a CLD. The CLD is thus defined by:

$$CLD = (DY_i, DCr_j, DCb_j)_{i=1...m, j=1...n} \tag{3}$$

Where DY_i represents the $i^{th}DCT$ coefficient of the Y color component, DCr_j represents the $j^{th}DCT$ coefficient of the Cr color component and DCb_j represents the $j^{th}DCT$ coefficient of the Cb color component. We note that m and n are defined by the user.

In our work, we take m=6 and n=3. So the CLD is represented by 12 values.

c) Scalable Color Descriptor (SCD): One of the most basic description of color features is provided by describing color distribution in images. SCD is defined in the hue-saturationvalue (HSV) color space with fixed color space quantization, and uses the Haar transform encoding that consists of an addition and a soustraction operation, which are related to primitive low and high-pass filters. Summing pairs of adjacent histogram lines is equivalent to the calculation of a histogram with half number of bins. If this process is performed iteratively, the use of subsets of the coefficients in the Haar representation is equivalent to histograms of 128, 64, 32 bins, which are all calculated from the source histogram. The Haar transform based encoding facilitates a scalable representation of the description, as well as complexity scalability for feature extraction and matching procedures.

In our case, the scalable color descriptor is presented by 64 values equals to 0 or 1.

B. Supervised learning

In the literature, there are several techniques of supervised learning, each having its advantages and disadvantages. In our approach, we study the use of the support vector machines [6], the graphs of decision tree [16] and the neural networks [17].

Given a training set $D = \{x_i, y_i\}_{i=1}^{N}$ with input vectors $x_i = (x_i^1, ..., x_i^n)^T \in R^n$ and target labels, we use, for the support vector machines, a Radial Basis Function as kernel defined by the following formula:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \tag{4}$$

Actually, in order to use the RBF kernel, appropriate values for the kernel parameters $\mathcal{C}$ (penalty) and γ (kernel width) need to be determined. Intensive experiments were used to determine the optimal values for these parameters. ($\mathcal{C} = 1, \gamma = 0, 3$) give the maximum prediction accuracy on learning and test data set.

For graphs of decision tree, five data mining algorithms including ID3 [18], C4.5 [19], IMPROVED C4.5 [20], SIPINA [21] and CHAID [22] have been experimented.

In order to train a model based on an Artificial Neural Networks, we use a multilayer perceptron with the back-propagation learning algorithm, composed of an input layer, an output layer, and one hidden layer. To choose the best parameters for MLP, we calculate the variation of the global error rate according to the number of neurons. The best result was scored by using 40 neurons. The activation function for the neurons in the hidden layer and in the output layer are sigmoid function defined by the following formula :

$$f(x) = \frac{1}{1 - \exp(-x)} \tag{5}$$

IV. EXPERIMENTATIONS

Actually, we carried out a series of experiments, focusing on visual content-based features, we decide to use several data mining techniques to derive learnt model, from a learning data set. The quality of this model is evaluated using random error rate technique. The stability of the learnt model is further validated using cross-validation and bootstrapping techniques. For the random error rate technique, we have used three measures to know: *global error rate*, *a priori error rate* and *a posteriori error rate*. As global error rate is the complement of classification accuracy rate, while *a priori error rate* (respectively, *a posteriori error rate*) is the complement of the classical *recall rate* (respectively, *precision rate*).

During this series of experiments, seven data mining algorithms have been studied, including C4.5, ID3, Improved C4.5, SIPINA, CHAID, MLP and SVM. The following figure shows the individual error rates of these algorithms on learning data set.

As we can see in figure 2, the SVM and MLP algorithm echoed very similar performance on the feature vector, displaying a global error rate of less than 8% and only 4% for the best one (SVM). For the decision tree algorithms, C4.5 presents
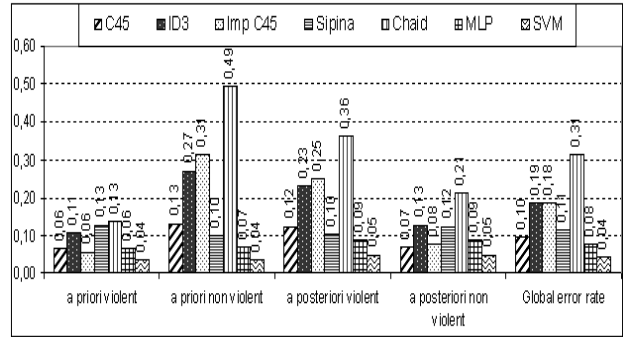

Fig. 2.   Experimental results by the seven algorithms on learning data set.

the best global error rate equal to 10%. The a priori error rates show that different algorithms are not as efficient as SVM concerning a priori error rate on violent Web images. This rate is important for evaluation because it quantifies the efficiency of the classification of violent Web images. Also, the same algorithm shows the best a priori non violent error rate. We can conclude that SVM produce the best learnt model for our problem.

A good predict model obtained by a data mining algorithm from the learning data set should not only produce good classification performance on data already seen but also on unseen data as well. In order to ensure the performance stability of our learnt model from learning data set and validated by random error rate technique, we have also tested the learnt model on our test data set. The experimental results are depicted in figure 3.
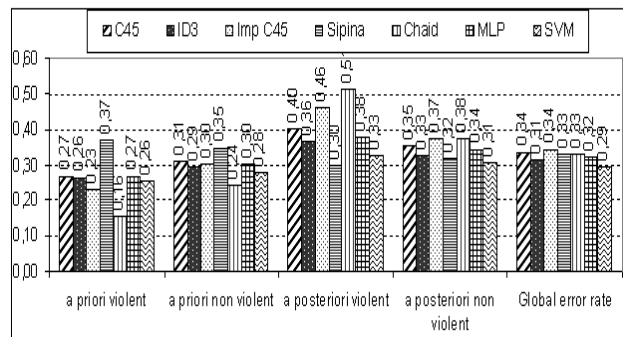

Fig. 3.   Experimental results by the seven algorithms on test data set.

All the seven data mining algorithms echoed very similar performance, displaying a global error rate of less than 34% and only 29% for the best one (SVM). There is clearly a tradeoff between a priori error and a posteriori rate on violent Web images, and the much the same for non violent Web images. For instance, when CHAID displayed the best performance on violent Web images with a priori rate of 24%, it recorded on the other hand a posteriori error rate of 51% which is the worst performance. The same tradeoff was also

observed on the non violent side, when CHAID achieved, with 24%, the best a priori error rate on non violent Web images, it recorded on the other hand a 38% of a posteriori error rate on non violent Web images which is the worst performance among the seven algorithms.

It seems that the best average behavior has been achieved by SVM, which records the best compromise between violent and non violent error rate. Taking into account both obtained results (learning and test), we choose it as the best algorithms. The experimental results by the random error rate technique are acceptable as all the seven algorithms. We therefore used the two other more systematic evaluation methods, bootstrapping [23] and cross-validation [24]to further confirm these results. Table I and Table II show the experimental results respectively by bootstrapping and cross-validation.

TABLE I
GLOBAL ERROR RATES FOR BOOTSTRAPPING.

	Bootstrapping
C45	0,3469
ID3	0,3416
Imp C4,5	0,3701
SIPINA	0,3546
CHAID	0,3385
SVM	0,2489
MLP	0,2719

The obtained results using the bootstrapping method confirm those obtained by random error rates. Indeed, SVM is the best classifier with a global error rate equal to 24,89%.

TABLE II
GLOBAL ERROR RATES FOR CROSS-VALIDATION.

	Cross-validation
C45	0,3172
ID3	0,2993
Imp C4,5	0,3536
SIPINA	0,3156
CHAID	0,3385
SVM	0,3171
MLP	0,3212

For the cross validation method, the best result was scored by ID3 (29,93%). Concerning SVM, even though the rate obtained with cross-validation method is the not the highest one, it does not differ from the best one by more that 0,179. Therefore, through the overall results of these tests, we can conclude that the results for these algorithms are acceptable and can been improved.

V. CLASSIFIER COMBINATION

As shown in figure 2 and figure 3, the five induction graphs algorithms (ID3, C4.5, SIPINA, Improved C4.5 and CHAID) displayed different performances on the various error rate. Although SVM gives the best results, combining algorithms may provide better results [14]. In our work, we have experimented two combining classifiers methods:

1) Scores combination: we thus affect a weight associated to the score classification decision of each data mining algorithm according to the following formula:

$$Score_{image} = \sum_{i=1}^{N} \alpha_i S_i \qquad (6)$$

Where α_i presents the belief value we accord to the algorithm i.

$$\alpha_i = \frac{\beta_i}{\sum_{j=1}^{N} \beta_i} \ with \ \beta_i = (1 - (\varepsilon_i - \delta))^n \qquad (7)$$

where

- ε_i: the global error rate of the i-th algorithm ;
- N : the number of algorithms used for classification. In our case N=5 ;
- n : the power in order to emphasize the difference in weight ;
- δ : a threshold value that we take away from the error rate again to emphasize the difference in weight.

To calculate the different weights, we are based on the global error rate in order to have the best balanced behavior of our classifier, on both violent and non violent classes. After several experiments, we have fixed n=5 and δ=0.085 giving the best results on test data set.

S_i presents the score to be a violent Web image according to i-th data mining algorithm $(i = 1 \dots 5)$. This combining method can be used only if all algorithms have same type. In our case we used five algorithms based on decision tree to know ID3, C4.5, SIPINA, Improved C4.5 and CHAID. The score is calculated for each algorithm as follows: having the decision tree, we seek for each image the node that corresponds to the values of its characteristics vector. Once found the score is calculated as the ratio of images judged as violent from the current node and its previous.

The image is considered as violent if its score is greater than a threshold which is determined automatically.

2) Learning the parameters of the combination function: The task of combining classifiers can be considered as a simple classification problem. the input vector contains the score of each individual decision tree, and the output is the forecast class. Indeed, it is possible to use classification algorithms to learn the parameters of the combination function. In our case, we have applied C4.5 to solve this combination problem.

Figure 4 presents a comparison between the scores combination model and the learning combination models on the learning data set. The experimental results by the random error rate technique are thus very encouraging as all the two combination methods with a priori error rate varying between 7% and 9% and a glabal error rate less than 13%.

Figure 5 presents a comparison between the scores combination model and the learning combination models on the test
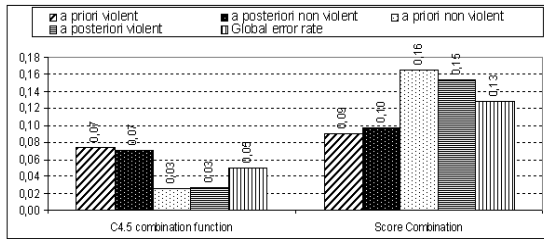
Fig. 4.   Comparison combination approaches on learning data set.
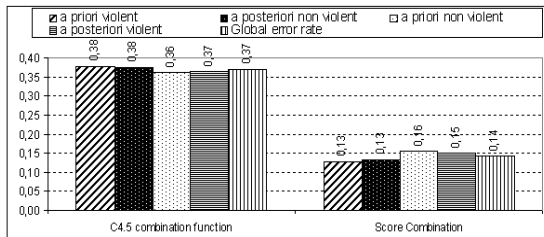

Fig. 5.   Comparison combination approaches on test data set.

data set.

The obtained results confirm the well interest to use the scores combination method. Indeed, it provides better results than C4.5 combination function. One clearly observes the reduction in the global error rate of 37% by the C4.5 combination method to 14% by the scores combination method. So, to build our violent Web image classifier, we are based on the predict model of the scores combination method.

VI. Conclusion

In this paper, we propose an approach of violent Web image classification based of the KDD process for extracting useful knowledge from volumes data. Such an approach depend strongly on the quality of the used features. In our work, we are based on MPEG 7 colors features to product predict models which allow to decide if an image is violent or not.

Our approach has shown its effectiveness, scoring a 86% classification accuracy rate on our test data set.

We can thus summarize our major contribution by : first, the study and comparison of several techniques of data mining, knowing, Support Vector Machines, Artificial Neural Networks and Decision tree to build a violent Web image classifier based on colors features. Second, combining classifiers to improve classification accuracy rate.

Our experimental evaluation shows the effectiveness of our approach, however, many future work directions can be considered. First, the introduction of other visual feature such as texture and shape. In fact, the color features used in our approach are still unable to cover all the visual apparence of the images. In addition, as we treat the Web images, we think that the exploitation of new textual features automatically extracted from the name, the url and the surrounding text image's with visual features can without any doubt ameliorate

the classification accuracy rate. Finally, integrate our approach of images classification in the violent Web sites classification.

References

[1] R. Guermazi, M. Hammami, and A. Ben-Hamadou, "Webangels filter: A violent web filtering engine using textual and structural content-based analysis," in *ICDM '08: Proceedings of the 8th industrial conference on Advances in Data Mining*.   Berlin, Heidelberg: Springer-Verlag, 2008, pp. 268–282.
[2] I. Sethi and I. Coman, "Mining association rules between low-level image features and high-level concepts," in *Proceedings of the SPIE Data Mining and Knowledge Discovery*, vol. III, 2001, pp. 279–290.
[3] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recogn.*, vol. 40, no. 1, pp. 262–282, 2007.
[4] J. Z. Wang, J. Li, and G. Wiederhold, "Simplicity: semantics- sensitive integrated matching for picture libraries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 9, pp. 947–963, 2001.
[5] R. Mehrotra and J. E. Gary, "Similar-shape retrieval in shape data management," *Computer*, vol. 28, no. 9, pp. 57–62, 1995.
[6] L. Zhang, F. Liu, and B. Zhang, "Support vector machine learning for image retrieval," in *International Conference on Image Processing*, October 2001, pp. 7–10.
[7] A. Vailaya, M. Figueiredo, A. Jain, and H. Zhang, "Image classification for content-based indexing," in *IEEE Transactions on Image Processing*, vol. 10, no. 1, 2001, pp. 117–130.
[8] C. Town and D. Sinclair, "Content-based image retrieval using semantic visual categories," Society for Manufacturing Engineers, Technical Report, Tech. Rep., 2001.
[9] S. MacArthur, C. Brodeley, and C.-R. Shyu, "Relevance feedback decision trees in content-based image retrieval," in *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries*, 2000, pp. 68–72.
[10] W.-C. Low and T.-S. Chua, "Color-based relevance feedback for image retrieval," in *Proceedings of the International Workshop on Multimedia DBMS*, 1998, pp. 116–123.
[11] Y. Chen, J. Wang, and R. Krovetz, "An unsupervised learning approach to content-based image retrieval," in *IEEE Proceedings of the International Symposium on Signal Processing and its Applications*, 2003, pp. 197–200.
[12] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)*, vol. 22, no. 8, pp. 888–905, 2000.
[13] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The kdd process for extracting useful knowledge from volumes of data," *Communications of the ACM (CACM)*, vol. 39, no. 11, pp. 27–34, 1996.
[14] R. Guermazi, M. Hammami, and A. Ben-Hamadou, "Combining classifiers for web violent content detection and filtering," in *International Conference on Computational Science*, vol. 3, 2007, pp. 772–779.
[15] N. Day and J. Martinez, "Introduction to mpeg-7 (v3.0)," 2001.
[16] D. A. Zighed and R. Rakotomalala, *Graphes d'Induction - Apprentissage et Data Mining*.   Hermes, 2000.
[17] J. Herault and C. Jutten, *Les rseaux de neurones et le traitement du signal*.   Herms, 1994.
[18] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, 1986.
[19] J.-R. Quinlan, *C4.5: Programs for Machine Learning*.   San Mateo: Morgan Kaufmann, 1993.
[20] R. Rakotomalala and S. Lallich, "Handling noise with generalized entropy of type beta in induction graphs algorithm," in *International Conference on Computer Science and Informatics*, 1998, pp. 25–27.
[21] D. A. Zighed and R. Rakotomalala, "A method for non arborescent induction graphs," Loboratory ERIC, University of Lyon 2, Technical report, 1996.
[22] G. V. Kass, "An exploratory technique for investigating large quantities of categorical data," *Applied Statistics*, vol. 29, pp. 119–127, 1980.
[23] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*.   Chapman & Hall, 1994.
[24] M. Stone, "Cross- validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society*, pp. 111–147, 1974.