

Color, texture and shape descriptor fusion with Bayesian network classifier for automatic image annotation

Mustapha OUJAOURA¹, Brahim MINAOU², Mohammed FAKIR³

Laboratory of Information Processing and Telecommunications, Computer Science Department,
Faculty of Science and Technology, Sultan Moulay Slimane University
Mghila, PO Box. 523, Béni Mellal, Morocco

Abstract—Due to the large amounts of multimedia data prevalent on the Web, Some images presents textural motifs while others may be recognized with colors or shapes of their content. The use of descriptors based on one's features extraction method, such as color or texture or shape, for automatic image annotation are not efficient in some situations or in absence of the chosen type. The proposed approach is to use a fusion of some efficient color, texture and shape descriptors with Bayesian networks classifier to allow automatic annotation of different image types. This document provides an automatic image annotation that merges some descriptors in a parallel manner to have a vector that represents the various types of image characteristics. This allows increasing the rate and accuracy of the annotation system. The Texture, color histograms, and Legendre moments, are used and merged respectively together in parallel as color, texture and shape features extraction methods, with Bayesian network classifier, to annotate the image content with the appropriate keywords. The accuracy of the proposed approach is supported by the good experimental results obtained from ETH-80 databases.

Keywords—image annotation; k-means segmentation; Bayesian networks; color histograms; Legendre moments; Texture; ETH-80 database

I. INTRODUCTION

With the rapid development of Internet communication technology and digital imaging technology, users can easily get many networked digital information archives by a variety of ways. Searching this digital information archives on the Internet and elsewhere has become a significant part of our daily lives. Amongst the rapidly growing body of information, many digital images are not reached. The task of automated image retrieval is complicated by the fact that many images do not have suitable textual descriptions and annotations. Retrieval of images through analysis of their visual content is therefore an exciting and notable research challenge.

With regard to the long standing problem of the semantic gap between low-level image features and high-level human knowledge, the image retrieval community has recently shifted its emphasis from low-level features analysis to high-level image semantics extraction. Therefore, image semantics extraction is of great importance to content-based image retrieval because it allows the users to freely express what images they want. Semantic content annotation is the basis for

semantic content retrieval. The automatically obtained keywords from image annotation process can be used to represent the images content to facilitate their retrieval.

Automatic object recognition and annotation are essential tasks in these image retrieval systems. Indeed, Annotated images play important role in information processing; they are useful for image retrieval based on keywords and image content management [1]. For that reason, many research efforts have aimed at annotating objects contained in visual streams. Image content annotation facilitates conceptual image indexing and categorization to assist text-based image search that can be semantically more significant than search in the absence of any text [2], [3].

Manual annotation is not only boring but also not practical in many cases, due to the abundance of information. Many images are therefore available without suitable textual annotation. Automatic image content annotation becomes a recent research interest [3], [4]. It attempts to explore the visual characteristics of images and associate them with image contents and semantics to use textual request for image retrieval and searching; automatic image annotation is an efficient technology for improving the image retrieval.

The rest of the paper is organized as follows. Firstly, the section 2 presents the proposed annotation system. The Section 3 discusses the image segmentation while the section 4 presents a brief formulation of color histograms, Texture, and Legendre moments as features extraction method. The Section 5 is reserved for the annotation by the approach of image classification using a fusion of several descriptors that are the color histograms, Texture, and Legendre moments with Bayesian network classifier. The Section 6 presents the experimental results of the image annotation based on the proposed approach. Finally, in the last section, the main conclusion concerning the proposed approach is given in addition to the possible future works.

II. ANNOTATION SYSTEM

Automatic image annotation consists of associating, to each image, a group of words that describes the visual contents of the image without human intervention. This task has been, and still, the subject of many studies [5], [6], [7], [8], [9], [10]. Several ways are used to deal with the problem of automatic image annotation. A recent review on automatic image

annotation techniques is presented in [11]. Using machine learning methods from examples of annotated images, many automatic image annotation techniques aim to learn the relationship between keywords and visual features. The learned relationships are then used to assign keywords to non-annotated images.

As an improvement of the previous works [12, 13, 14], this problem is considered for image that are described globally or not yet labeled with a suitable text terms. Some images on the web presents textural motifs, others can be recognized with colors or shapes of their content. The fusion of multiple descriptors, which are of different types such as color or texture or shape descriptors, can increase the effectiveness of images representation allowing their annotations in a manner that is more accurate than when using one descriptor type's. Indeed, in situations where images have a different descriptor type's from that used, the results will be catastrophic. Also, individual classifier and features descriptor results are limited or not suited for some situations [15]. So, their fusion is important, it can improve the annotation results and improve the accuracy of the annotation system. In such case, the proposed approach is to use a fusion of some efficient color, texture and shape descriptors with Bayesian network classifier to allow automatic annotation of different image types. The objective of this document is to provide an automatic image annotation that merges some descriptors in a parallel manner to have a features vector that represents the various types of image characteristics. This approach can allow increasing the rate and accuracy of the annotation system. The block diagram of the image annotation system adopted in this work is shown in Fig.1.

The system contains several phases. Firstly, the query image is segmented into regions that represent objects in the image, secondly, the features vector of each region are computed and extracted from the image, and those features are merged and are finally fed into input of the already trained classifiers that is the Bayesian Network to decide and choose the appropriate keywords for annotation tasks.

III. K-MEANS IMAGE SEGMENTATION

Usually, the features vector extracted from the entire image loses local information. Therefore, it is necessary to segment an image into regions or objects of interest and use of local characteristics. Image segmentation is a method that localizes and extracts an object from an image or divides the image into several regions. It plays important role in many applications for image processing, and still remains a challenge for scientists and researchers.

The efforts and attempts are still being made to improve the segmentation techniques. With the improvement of computer processing capabilities, several possible segmentation techniques of an image have emerged: threshold, region growing, k-means, active contours, level sets, etc...[16].

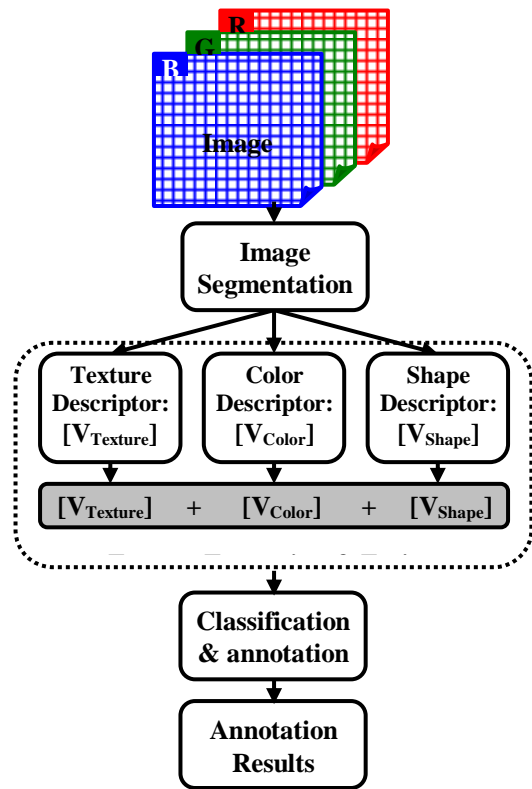


Fig. 1. Block Diagram of the proposed annotation system.

Among the segmentation methods, the k-means is well suited because of its simplicity and has been successfully used several times as a segmentation technique of digital images.

The K-Means algorithm is based on a clustering algorithm that does not require the presence of a learning database. So, this algorithm can organize the pixels of the image.

Given a set of image pixels $X = \{p_1, p_2, \dots, p_n\} \in R^d$ where each pixel is a veritable vector of dimension $d = 3$ in the case of a colour image ($d = 5$ if the pixels coordinates are introduced as information of spatial coherence or connectivity). The k-Means algorithm aims to classify and divide the n pixels of the image into k sets or regions ($k \leq n$) $S = \{R_1, R_2, \dots, R_k\}$ with a manner that minimizes the inter-class variance, that results in minimizing the sum of squared Euclidean distances among the clusters defined by:

$$E = \sum_{i=1}^k \sum_{p_j \in R_i} \|p_j - m_i\|^2 = \sum_{i=1}^k \text{Card}(R_i) \times \text{Var}(R_i) \quad (1)$$

Where:

- p_j pixel vector;

- $Card(R_i)$ is the number of pixels in the cluster or region R_i ;

- $m_i = \frac{\sum_{p_j \in R_i} p_j}{Card(R_i)}$ is the centre of the cluster or region R_i ; also known as kernel;

- $Var(R_i) = \frac{\sum_{p_j \in R_i} \|p_j - m_i\|^2}{(Card(R_i))^2}$ is the variance of pixel cluster or region.

The k-means image segmentation algorithm finds the pixels groups that minimize the quantity E defined above. This comes somehow for each cluster or region, to minimize the following quantity:

$$\sum_{p_j \in R_i} \|p_j - m_i\|^2 \quad (2)$$

The principle of the minimization algorithm of this error can result in the following main steps [18]:

- 1) Choosing the number of clusters (number of kernels);
- 2) Initialization of clusters and their kernels;
- 3) Updating clusters by optimizing the error clustering;
- 4) Calculation and revaluation of the new clusters;
- 5) Iterate and repeat steps 3 and 4 until clusters stabilization.

The number of clusters k can match approximately the number of dominant colors used to represent the image. The determination of k is done using the color histograms.

After transformation of the color image into a single image formed by the reduced numbers of colors, the cluster number k is selected to be the number of peaks in the histogram from the transformed image.

An example of image segmentation, by using K-means segmentation algorithm, is presented in Fig. 2.

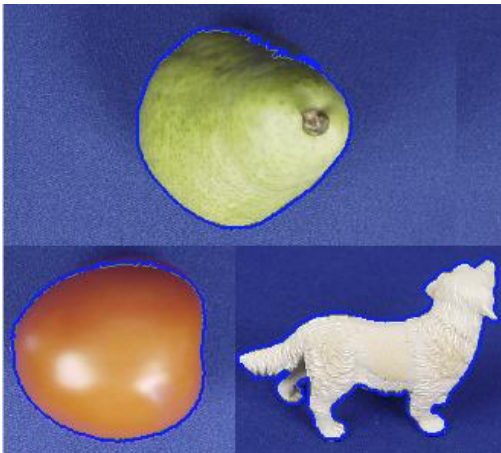


Fig. 2. Example of K-means image segmentation.

IV. FEATURES EXTRACTION

After dividing the original image into several distinct regions that correspond to objects in a scene, the feature vector must be extracted carefully from a region to reduce the rich content and large input data of images and preserve the content representation of the entire image. Therefore, the feature extraction task can decrease the processing time. It enhances not only the retrieval and annotation accuracy, but also the annotation speed as well, since a large image database can be organized according to the classification rule and, therefore, search can be done [19].

In the feature extraction method, the representation of the image content must be considered in some situations such as: translation, rotation and change of scale. This is the reason that justifies the use of color histograms and moments for feature extraction method from the segmented image.

All these features are extracted for all the images in reference database and stored with keywords in features database. For more precision and accuracy in the annotation system, they can be combined together and feed to the input of the classifier [15]. This combination costs more time for training the classifiers due to the size of the resulted features.

A. Color histogram

Typically, the color of an image is represented through some color model. There exist various color models to describe color information. The more commonly used color models are RGB (red, green, blue), HSV (hue, saturation, value) and Y, Cb, Cr (luminance and chrominance). Thus, the color content is characterized by 3 channels from some color models. In this paper, we used RGB color models. One representation of color image content is by using color histogram. Statistically, it denotes the joint probability of the intensities of the three color channels [20].

Color histogram describes the distribution of colors within a whole or within an interest region of image. The histogram is invariant to rotation, translation and scaling of an object but the histogram does not contain semantic information, and two images with similar color histograms can possess different contents.

The histograms are normally divided into bins to coarsely represent the content and reduce dimensionality of subsequent classification and matching phase. A color histogram H for a given image is defined as a vector by:

$$H = \left\{ h[i \in \{1, \dots, k\}] = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \delta(f(x, y) - C(i))}{M \times N} \right\} \quad (3)$$

$$\left\{ / (i-1) \times E\left(\frac{256}{k}\right) \leq C(i) < i \times E\left(\frac{256}{k}\right) \right\}$$

Where:

- i represent a color in the color histogram;
- E(x) denotes the integer part of x;

- $h[i]$ is the number of pixel with color i in that image;
- k is the number of bins in the adopted color model;
- And δ is the unit pulse defined by:

$$\delta(x, y) = \begin{cases} 1 & \text{if } x = y = 0 \\ 0 & \text{else} \end{cases} \quad (4)$$

In order to be invariant to scaling change of objects in images of different sizes, color histograms H should be divided by the total number of pixels $M \times N$ of an image to have the normalized color histograms.

For a three-channel image, a feature vector is then formed by concatenating the three channel histograms into one vector.

B. Legendre Moments

In this paper, the Legendre moments are calculated for each one of the 3 channel in a color image. A feature vector is then formed by concatenating the three channel moments into one vector.

The Legendre moments [21] for a discrete image of $M \times N$ pixels with intensity function $f(x, y)$ is the following:

$$L_{pq} = \lambda_{pq} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} P_p(x_i) P_q(y_j) f(x, y) \quad (5)$$

Where $\lambda_{pq} = \frac{(2p+1)(2q+1)}{M \times N}$, x_i and y_j denote the normalized pixel coordinates in the range of $[-1, +1]$, which are given by:

$$\begin{cases} x_i = \frac{2x - (M-1)}{M-1} \\ y_j = \frac{2y - (N-1)}{N-1} \end{cases} \quad (6)$$

$P_p(x)$ is the p^{th} -order Legendre polynomial defined by:

$$P_p(x) = \sum_{k=0}^p \left\{ \frac{(-1)^{\frac{p-k}{2}} (p+k)! x^k}{2^p k! \left(\frac{p-k}{2}\right)! \left(\frac{p+k}{2}\right)!} \right\}_{p-k=\text{even}} \quad (7)$$

In order to increase the computation speed for calculating Legendre polynomials, we used the recurrent formula of the Legendre polynomials defined by:

$$\begin{cases} P_p(x) = \frac{(2p-1)x}{p} P_{p-1}(x) - \frac{(p-1)}{p} P_{p-2}(x) \\ P_1(x) = x, \quad P_0(x) = 1 \end{cases} \quad (8)$$

C. Texture Descriptors

Several images have textured patterns. Therefore, the texture descriptor is used as feature extraction method from the segmented image.

The texture descriptor is extracted using the co-occurrence matrix introduced by Haralick in 1973 [22]. So for a color image I of size $N \times N \times 3$ in a color space (C_1, C_2, C_3) , for $(k, l) \in [1, \dots, N]^2$ and $(a, b) \in [1, \dots, G]^2$, the co-occurrence matrix $M_{k,l}^{C,C'}[I]$ of the two color components $C, C' \in \{C_1, C_2, C_3\}$ from the image I is defined by:

$$M_{k,l}^{C,C'}([I], a, b) = \frac{1}{(N-k)(N-l)} \times \sum_{i=1}^{N-k} \sum_{j=1}^{N-l} \delta(I(i, j, C) - a, I(i+k, j+l, C') - b) \quad (9)$$

Where δ is the unit pulse defined by:

$$\delta(x, y) = \begin{cases} 1 & \text{if } x = y = 0 \\ 0 & \text{else} \end{cases} \quad (10)$$

Each image I in a color space (C_1, C_2, C_3) can be characterized by six color co-occurrence matrix.

Matrix $M^{C_2, C_1}[I]$, $M^{C_3, C_1}[I]$ and $M^{C_3, C_2}[I]$ are not taken into account because they can be deduced respectively by diagonal symmetry from matrix $M^{C_1, C_2}[I]$, $M^{C_1, C_3}[I]$ and $M^{C_2, C_3}[I]$. As they measure local interactions between pixels, they are sensitive to significant differences in spatial resolution between the images. To reduce this sensitivity, it is necessary to normalize these matrices by the total number of the considered co-occurrences matrix:

$$M_{k,l}^{C,C'}([I], a, b) = \frac{M_{k,l}^{C,C'}([I], a, b)}{\sum_{i=0}^{T-1} \sum_{j=0}^{T-1} M_{k,l}^{C,C'}([I], i, j)} \quad (11)$$

Where T is the number of quantization levels of the color components.

To reduce the large amount of information of these matrices, the 14 Haralick indices [22] of these matrices are used. There will be then 84 textures attributes for six co-occurrence matrices (14×6).

V. IMAGE CLASSIFICATION AND ANNOTATION

The goal of pattern classification is to allocate an object represented by a number of feature vectors into one of a finite set of classes from the reference database. In order to classify unknown patterns, a certain number of training samples available for each class are used to train the classifier. The learning task is to compute a classifier or model that approximates the mapping between the input-output examples and correctly labels the training set with some level of accuracy. This can be called the training or model generation stage. After the model is generated and trained, it is able to classify an unknown instance, into one of the learned class labels in the training set. More specifically, the classifier calculates the similarity of all trained classes and assigns the unlabeled instance to the class with the highest similarity measure.

Therefore, image annotation can be approached by the model or the classifier generated and trained to bridge the gap between low-level feature vectors and high-level concepts; a function is learned which can directly correspond the low-level feature sets to high-level conceptual classes. There are several types of classifier that can be used for classification. The Bayesian network classifier is used in this paper.

Bayesian networks are based on a probabilistic approach governed by Bayes' rule. The Bayesian approach is then based on the conditional probability that estimates the probability of occurrence of an event assuming that another event is verified. A Bayesian network is a graphical probabilistic model representing the random variable as a directed acyclic graph. It is defined by [22]:

- $G = (X, E)$, Where X is the set of nodes and E is the set of edges, G is a Directed Acyclic Graph (DAG) whose vertices are associated with a set of random variables $X = \{X_1, X_2, \dots, X_n\}$;
- $\theta = \{P(X_i | Pa(X_i))\}$ is a conditional probabilities of each node X_i relative to the state of his parents $Pa(X_i)$ in G.

The graphical part of the Bayesian network indicates the dependencies between variables and gives a visual representation tool of knowledge more easily understandable by users. Bayesian networks combine qualitative part that are graphs and a quantitative part representing the conditional probabilities associated with each node of the graph with respect to parents [23].

Pearl and all [24] have also shown that Bayesian networks allow to compactly representing the joint probability distribution over all the variables:

$$P(X) = P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i)) \quad (12)$$

Where $Pa(X_i)$ is the set of parents of node X_i in the graph G of the Bayesian network.

This joint probability could be actually simplified by the Bayes rule as follows [25]:

$$\begin{aligned} P(X) &= P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i)) \\ &= P(X_n | X_{n-1}, \dots, X_1) \times P(X_{n-1} | X_{n-2}, \dots, X_1) \times \dots \times P(X_2 | X_1) \times P(X_1) \\ &= P(X_1) \times \prod_{i=2}^n P(X_i | X_{i-1}, \dots, X_1) \end{aligned} \quad (13)$$

The construction of a Bayesian network consists in finding a structure or a graph and estimates its parameters by machine learning. In the case of the classification, the Bayesian network can have a class node C_i and many attribute nodes X_j . The naive Bayes classifier is used in this paper due to its robustness and simplicity. The Fig. 3 illustrates its graphical structure.

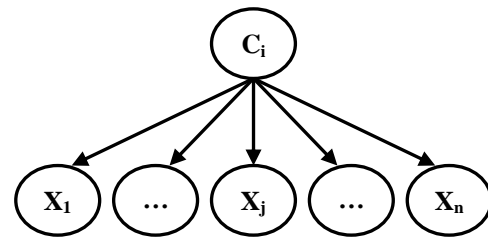


Fig. 3. Naive Bayes classifier structure.

To estimate the Bayesian network parameters and probabilities, Gaussian distributions are generally used. The conditional distribution of a node relative to its parent is a Gaussian distribution whose mean is a linear combination of the parent's value and whose variance is independent of the parent's value [26]:

$$P(X_i | Pa(X_i)) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left\{ \frac{-1}{2\sigma_i^2} \left(x_i - \left(\mu_i + \sum_{j=1}^{n_i} \frac{\sigma_{ij}}{\sigma_j^2} (x_j - \mu_j) \right) \right)^2 \right\} \quad (14)$$

Where,

- $Pa(X_i)$ Are the parents of X_i ;
- μ_i, μ_j, σ_i and σ_j are the means and variances of the attributes X_i and X_j respectively without considering their parents;
- n_i is the number of parents;

- σ_{i_j} is the regression matrix of weights.

After the parameter and structure learning of a Bayesian network, The Bayesian inference is used to calculate the probability of any variable in a probabilistic model from the observation of one or more other variables. So, the chosen class C_i is the one that maximizes these probabilities [27], [28]:

$$P(C_i|X) = \begin{cases} P(C_i) \prod_{j=1}^n P(X_j | Pa(X_j), C_i) & \text{if } X_j \text{ has parents} \\ P(C_i) \prod_{j=1}^n P(X_j | C_i) & \text{else} \end{cases} \quad (15)$$

For the naive Bayes classifier, the absence of parents and the variables independence assumption are used to write the posterior probability of each class as given in the following equation [29]:

$$P(C_i|X) = P(C_i) \prod_{j=1}^n P(X_j | C_i) \quad (16)$$

Therefore, the decision rule d of an attribute X is given by:

$$\begin{aligned} d(X) &= \arg \max_{C_i} P(C_i|X) \\ &= \arg \max_{C_i} P(X|C_i) P(C_i) \\ &= \arg \max_{C_i} P(C_i) \prod_{j=1}^n P(X_j | C_i) \end{aligned} \quad (17)$$

The class with maximum probability leads to the suitable character for the input image.

VI. EXPERIMENTS AND RESULTS

A. Experiments

In the experiments, for each region that represent an object from each color channel of the query image, the number of input features extracted using the order 3 of Legendre moments is 10 (L00, L01, L02, L03, L10, L11, L12, L20, L21, L30). The number of input features for color histogram is 16 per image channel. So, the result is 30 elements for Legendre moments and 48 elements for color histograms in the case of the 3 channels. The number of input features extracted using Texture extraction method is $14 \times 6 = 84$. These inputs are presented and feed to the Bayesian network classifier, for testing to do matching with the feature values in the reference database. To test the accuracy of the proposed approach, we measured the precision rates of each single descriptor.

Fig. 4 shows some examples of image objects from ETH-80 image database used in our experiments. The experiments are made based on different classes of objects.



Fig. 4. Some examples of objects from ETH-80 image database.

The accuracy of image annotation is evaluated by the precision rate which is the number of correct results divided by the number of all returned results.

All the experiments are conducted using the ETH-80 database containing a set of 8 different object images [30]. The proposed system has been implemented and tested on a core 2 Duo personnel computer using Matlab software.

B. Results

The results of the image annotation system based on Legendre, RGB, Texture descriptors and their fusion using a Bayesian network classifier are presented in table 1.

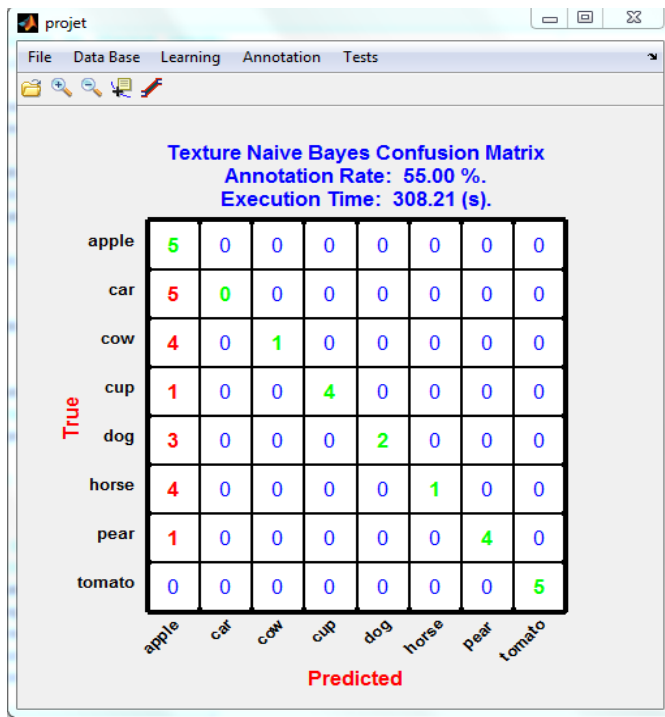
TABLE I. GENERAL ANNOTATION RATES OF THE ANNOTATION SYSTEM BASED ON LEGENDRE, RGB, TEXTURE DESCRIPTORS AND THEIR FUSION USING A BAYESIAN NETWORK CLASSIFIER.

Descriptor Approach	Annotation rate (%)	Error rate (%)	Execution time (s)
Legendre	78.00%	22.00%	10756.49
RGB	67.50%	32.50%	162.16
Texture	55.00%	45.00%	489.65
Legendre+Texture+RGB+Bayes	87.50%	12.50%	12430.78

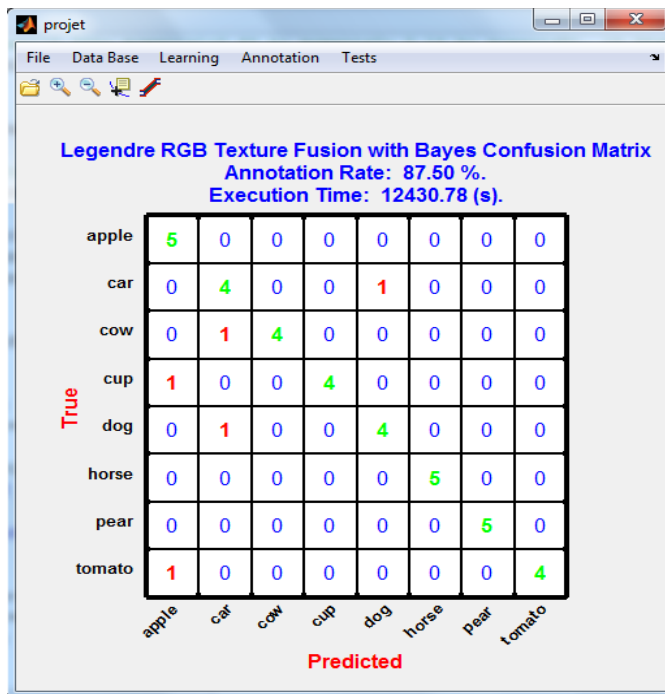
From the Table 1, the experimental results showed that the annotation rate of the proposed method based on the fusion of Texture, color histogram and shape descriptor increase the precision of the color image annotation system.

The fusion of descriptors will certainly increase significantly the annotation rate since they will fill each other. In some situation where one or two descriptors are not suitable, the 2nd or the other descriptors can give a good result. The object in the input image will be recognized and annotated by either the texture or the color histogram or the shape descriptors.

In order to show the robustness of the proposed approach, more details are provided by calculating the confusion matrix of some method as presented in Fig. 5.



a)



b)

Fig. 5. Confusion matrix for: a) Texture with bayesian network, b) Texture, color and shape descriptors fusion with bayesian network.

The 2 confusion matrix in Fig. 5 show that the misclassified and incorrectly annotated objects (indicated by red color) in the case of using texture as single descriptor and Bayesian network classifier are reduced in the case of using the proposed approach based on fusion of many descriptors kinds (Texture, color and shape).

The results are also affected by the accuracy of the image segmentation method. In most cases, it is very difficult to have an automatic ideal segmentation. Therefore, any annotation attempt must consider image segmentation as an important step, not only for automatic image annotation system, but also for the other systems which requires its use.

VII. CONCLUSION

In this paper, the approach based on a fusion of different descriptors is used for the automatic image annotation system. For this image annotation system, we discussed the effect of merging different type of descriptors. The texture, color histogram and shape descriptors are merged together in one feature vector and used to classify and annotate the input image by the suited keywords that are selected from the reference database image. The performance of the proposed method has been experimentally analyzed. The successful experimental results proved that the proposed image annotation system gives good results for image that are well and properly segmented. However, Image segmentation remains a challenge that needs more attention in order to increase precision and accuracy of the image annotation system. Also, the gap between the low-level features and the semantic content of an image must be reduced and considered for more accuracy of any image annotation system. Other segmentation method and other features extraction method must be considered for future work. The feedback of results can be investigated for the automatic image annotation system. Finally, the execution time must be decreased in order to use the online system.

REFERENCES

- [1] Datta, R., Joshi, D., Li, J., and Wang, J. Z. 2008. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.* 40, 2, Article 5 (April 2008), 60 pages DOI = 10.1145/1348246.1348248 <http://doi.acm.org/10.1145/1348246.1348248>.
- [2] N. Vasconcelos and M. Kunt, *Content-Based Retrieval from Image Databases: Current Solutions and Future Directions*, Proc. Int'l Conf. Image Processing, 2001.
- [3] Shao, W, Naghdy, G and Phung, SL, *Automatic Image Annotation for Semantic Image Retrieval*, Lecture Notes in Computer Science, 4781, 2007, 369-378. Copyright Springer-Verlag.
- [4] Lei Ye, Philip Ogunbona and Jianqiang Wang, *Image Content Annotation Based on Visual Features*, Proceedings of the Eighth IEEE International Symposium on Multimedia (ISM'06), IEEE computer society, San Diego, USA, 11-13 December 2006.
- [5] J. P. Fan, Y. Gao & H. Z. Luo. Integrating Concept Ontology and Multitask Learning to Achieve More Effective Classifier Training for Multilevel Image Annotation. *IEEE Trans. Image Processing*, vol. 17, no. 3, pages 407–426, 2008.
- [6] X. J. Wang, L. Zhang, X. Li & W. Y. Ma. Annotating Images by Mining Image Search Results. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pages 1919–1932, 2008.
- [7] R. C. F. Wong & C. H. C. Leung. Automatic Semantic Annotation of Real-World Web Images. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pages 1933–1944, 2008.
- [8] J.Z Wang, Jia Li. Real-Time Computerized Annotation of Pictures, *Pattern Analysis and Machine Intelligence*, IEEE Transactions on Volume: 30, Issue: 6, pp. 985-1002, 2008
- [9] Ballan, Lamberto; Bertini, Marco; Uricchio, Tiberio; Del Bimbo, Alberto, *Social media annotation*, 11th International Workshop on Content-Based Multimedia Indexing 2013 (CBMI), pp. 229-235, 17-19 June 2013.
- [10] Qi Mao; Tsang, I.W.-H.; Shenghua Gao, "Objective-Guided Image Annotation," *Image Processing*, IEEE Transactions on , vol.22, no.4, pp.1585-1597, April 2013.

- [11] Zhang, D., Islam, M. M., and Lu, G., A review on automatic image annotation techniques. *Pattern Recognition*, 45(1): pp. 346 – 362, 2012.
- [12] Mustapha Oujaoura, Brahim Minaoui and Mohammed Fakir. Article: Image Annotation using Moments and Multilayer Neural Networks. *IJCA Special Issue on Software Engineering, Databases and Expert Systems SEDEX (1):46-55*, September 2012. Published by Foundation of Computer Science, New York, USA.
- [13] Mustapha Oujaoura, Brahim Minaoui and Mohammed Fakir, Multilayer Neural Networks and Nearest Neighbor Classifier Performances for Image Annotation, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 11, pp.165–171, 2012. Published by The Science and Information Organization, New York, USA.
- [14] Oujaoura, M.; Minaoui, B.; Fakir, M., "A semantic approach for automatic image annotation," *Intelligent Systems: Theories and Applications (SITA)*, 2013 8th International Conference on , vol., no., pp.1–8, 8-9 May 2013, doi: 10.1109/SITA.2013.6560800.
- [15] R. Sinan Tumen, M. Emre Acer and T. Metin Sezgin, Feature Extraction and Classifier Combination for Image-based Sketch Recognition, *Eurographics Symposium on Sketch-Based Interfaces and Modeling*, pp. 1–8, 2010.
- [16] Frank Y. Shih, Shouxian Cheng, Automatic seeded region growing for color image segmentation, *Image and Vision Computing* 23, pp. 877–886, 2005.
- [17] Aristidis Likas, Nikos A. Vlassis, and Jakob J. Verbeek. The global k-means clustering algorithm. *Pattern Recognition*, 36(2): pp. 451–461, 2003.
- [18] Lior Rokach, Oded Maimon. *Data Mining and Knowledge Discovery Handbook*, Chapter 15: Clustering Methods. pp 321-352, Springer series, 2nd Edition, New York, October 1, 2010.
- [19] Zijun Yang and C.-C. Jay Kuo, Survey on Image Content Analysis, Indexing, and Retrieval Techniques and Status Report of MPEG-7, *Tamkang Journal of Science and Engineering*, Vol. 2, No. 3, pp. 101-118, 1999.
- [20] Ryszard S. Chora's. Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems, *International Journal Of Biology And Biomedical Engineering*, Issue 1, Vol. 1, pp. 6-16, 2007.
- [21] Chee-Way Chonga, P. Raveendranb and R. Mukundan, Translation and scale invariants of Legendre moments, *Pattern Recognition* 37, pp. 119 – 129, 2004.
- [22] R. Haralick, K. Shanmugan, et I. Dinstein. Textural features for image classification. *IEEE Transactions on SMC*, 3(6) : pp. 610–621, 1973.
- [23] Ann.Becker, Patrick Naim : les réseaux bayésiens : modèles graphiques de connaissance. Eyrolles.1999.
- [24] J. Pearl, "Bayesian Networks" UCLA Cognitive Systems Laboratory, Technical Report (R-216), Revision I. In M. Arbib (Ed.), *Handbook of Brain Theory and Neural Networks*, MIT Press, 149-153, 1995.
- [25] Sabine Barrat, Modèles graphiques probabilistes pour la reconnaissance de formes, thèse de l'université Nancy 2, Spécialité informatique, décembre 2009.
- [26] George H. John and Pat Langley. Estimating continuous distributions in bayesian classifiers, the Eleventh Conference on Uncertainty in Artificial Intelligence, 1995.
- [27] Philippe LERAY, Réseaux bayésiens : apprentissage et modélisation de systèmes complexes, Habilitation A Diriger Les Recherches, Spécialité Informatique, Automatique et Traitement du Signal, Université de Rouen, novembre 2006.
- [28] Patrick Naïm, Pierre Henri Wuillemin, Philippe Leray, Olivier pourret, Anna becker, Réseaux bayésiens, Eyrolles, 3ème édition, Paris, 2008.
- [29] Tom .Mitchell: Generative and discriminative classifier: Naïve bayes and logistic regression. *Machine learning*. Draft 2010.
- [30] ETH-80 database image. [Online]. Available: <http://www.d2.mpi-inf.mpg.de/Datasets/ETH80>