DOI: 10.11591/telkomnika.v15i3.8536

# Toward an Effective Combination of Multiple Visual Features for Semantic Image Annotation

**B. Minaoui*[1], M. Oujaoura[2], M. Fakir[3], M. Sajieddine[4]**
[1,2,3]Laboratory of Information Processing & Decision
[4]Material Physics Laboratory
Faculty of Science and Technology, Sultan Moulay Slimane University, Mghila,
PO Box.523, Beni Mellal, Morocco.
*Corresponding author, email: bra_min@yahoo.fr

***Abstract***

*In this paper we study the problem of combining low-level visual features for semantic image annotation. The problem is tackled with a two different approaches that combines texture, color and shape features via a Bayesian network classifier. In first approach, vector concatenation has been applied to combine the three low-level visual features. All three descriptors are normalized and merged into a unique vector used with single classifier. In the second approach, the three types of visual features are combined in parallel scheme via three classifiers. Each type of descriptors is used separately with single classifier. The experimental results show that the semantic image annotation accuracy is higher when the second approach is used.*

***Keywords****: combining, fusion, features, image, annotation, semantic, bayesian networks, classifier*

## 1. Introduction

With the emergence of multimedia technology and rapid increase in the size of digital image collections, developing effective and efficient image retrieval systems becomes a very active research area. Several image retrieval techniques has been realized and extensively studied. It has been shown that the semantic gap between low-level image features and high-level semantic concepts is still the key hindrance in the effectiveness of these techniques. How to bridge the gap between visual features and semantic features has been a major challenge in this research area. The research conducted, in this regard, has widely recognized that the typical method of bridging the semantic gap is the automatic image annotation which is a process that automatically assigning, from a vocabulary of semantic concepts, a set of keywords (labels) to a digital image. Most of automatic image annotation approaches developed consist to learn semantic concept models that extract low-level visual features representing visual information of an image, and associate them automatically to high-level semantic concepts from a large number of training image samples, and use them to label new images.

Recently, in order to improve the quality of image annotation, a number research efforts [1-13] has focused on the problem of how to extract, from low-level features, the semantic concepts that interprets well the contents of image (objects, themes, events).

They have shown that the image annotation techniques should integrate different types of low-level visual features of visual data. None of the feature descriptors is sufficient to tackle intra-class diversity and inter-class correlation in an effective or efficient way. Therefore, effective low level features combination for image annotation and retrieval has become a desirable and promising perspective from which the semantic gap can be further bridged.

Motivated by this fact, we are interested to search a way to efficiently combine the strength of diverse and complementary features for semantic image annotation.

In this work, in an attempt to achieve this goal, we are interested in studying two approaches for combining three different kind of visual features (including color, texture and shape descriptors) for semantic image annotation. To evaluate the performance of these approaches, we have used Bayesian networks classifier and three image databases ETH-80 [14], COL-100 [15] and NATURE [16].

*Received June 3, 2015; Revised July 15, 2015; Accepted August 10, 2015*

This paper is organized as follows: The various low level visual features used in this study are presented in Section II. Section III presents the Bayesian networks classifier. Section IV describes the experiences adopted to realize the semantic image annotation using different combinations of low level visual features presented in section II. Finally, the conclusion of this work is presented in Section V.

## 2. Low Level Visual Features

After dividing the original image into several distinct regions that correspond to objects in a scene by using region growing segmentation algorithm [17], the color histogram, Haralick texture and Legendre moments descriptors are extracted.

The selection of these types of features is based on the consideration that they have proven discriminative and complementary for representing images capturing the semantics in the general scope of real life.

### 2.1. 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.

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.

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:

$$H = \left\{ h[i \in \{1,...,k\}] = \frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} \text{¢}\left(f(x,y) - C(i)\right)}{M \times N} \quad where: (i-1) \times E\left(\frac{256}{k}\right) \le C(i) < i \times E\left(\frac{256}{k}\right) \right\} \tag{1}$$

Where:
f(x, y) is the intensity function
i represent a color in the color histogram;
E(x) denotes the integer part of x;
h[i] is the number of pixel with colori in that image;
k is the number of bins in the adopted color model;
And δ is the unit pulse defined by:

$$\text{¢}(x\,,\,y) = \begin{cases} 1 & if \quad x = y = 0 \\ 0 & else \end{cases} \tag{2}$$

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 x N of an image in order to have the normalized color histograms.

For a three-channel image, we will have three of such histograms. A feature vector is then formed by concatenating the three channel histograms into one vector.

### 2.2. 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 [18] for a discrete image of M x N pixels with intensity function f(x, y) is the following:

$$L_{pq} = \text{a}_{pq} \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} P_p(x_i)\, P_q(y_j)\, f(x,y) \tag{3}$$

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

$$\begin{cases} x_i = \frac{2x-(M-1)}{M-1} \\ y_j = \frac{2y-(N-1)}{N-1} \end{cases} \tag{4}$$

$P_p(x)$ is the pth-order Legendre polynomial defined by:

$$P_p(x) = \sum_{k=0}^{p} \left\{ \frac{(-1)^{\frac{p-k}{2}} (p+k)!\, x^k}{2^p\, k! \left(\frac{p-k}{2}\right)! \left(\frac{p+k}{2}\right)!} \right\}_{p-k=even} \tag{5}$$

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

$$\begin{cases} P_p(x) = \frac{(2p-1)x}{p}\, P_{p-1}(x) - \frac{(p-1)}{p} P_{p-2}(x) \\ P_1(x) = x \quad , \quad P_0(x) = 1 \end{cases} \tag{6}$$

### 2.3. Haralick Texture Features

The texture descriptor is extracted using the co-occurrence matrix introduced by Haralick in 1973 [19]. So for a color image I of size $N \times N \times 3$ in a color space $(C_1, C_2, C_3)$, for $(k, l) \in [1, \cdots, N]^2$ and $(a, b) \in [1, \cdots, G]^2$, 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:

$$M_{k,l}^{C,C'}([I], a, b) = \frac{\sum_{i=1}^{N-k}\sum_{j=1}^{N-l} \text{¢}\left(I(i,j,C)-a\, ,\, I(i+k, j+l, C')-b\right)}{(N-k)(N-l)} \tag{7}$$

Where δ is the unit pulse defined by:

$$\text{¢}(x\, ,\, y) = \begin{cases} 1 & if \quad x = y = 0 \\ 0 & else \end{cases} \tag{8}$$

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

$$M^{C_1,C_1}[I],\ M^{C_2,C_2}[I],\ M^{C_3,C_3}[I],\ M^{C_1,C_2}[I],\ M^{C_1,C_3}[I], M^{C_2,C_3}[I]$$

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:

$$M_{k,l}^{C,C'}([I],a,b)=\frac{M_{k,l}^{C,C'}([I],a,b)}{\sum_{i=0}^{T-1}\sum_{j=0}^{T-1}M_{k,l}^{C,C'}([I],i,j)} \tag{9}$$

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 [16] of these matrices are used. There will be then 84 textures attributes for six co-occurrence matrices $(14\times 6)$.

## 3. Bayesian Networks Classifier

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 networks is a graphical probabilistic model representing the random variable as a directed acyclic graph. It is defined by [20]:

a) $G=(X\ ,E)$, 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 $X=\{X_1,X_2,\cdots,X_n\}$;

b) $^{-}=\{P(X_i|Pa(X_i))\}$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 is graphs and a quantitative part representing the conditional probabilities associated with each node of the graph with respect to parents.

Pearl and all [21] have also shown that Bayesian networks allow to compactly representing the joint probability distribution over all the variables:

$$P(X)=P(X_1,X_2,\cdots,X_n)=\prod_{i=1}^{n}P(X_i|Pa(X_i)) \tag{10}$$

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

This joint probability could be actually simplified by the Bayes rule as follows [22]:

$$\begin{aligned}P(X)&=P(X_1,X_2,\cdots,X_n)=\prod_{i=1}^{n}P(X_i|Pa(X_i))\\&=P(X_n|X_{n-1},\cdots,X_1)\times P(X_{n-1}|X_{n-2},\cdots,X_1)\times\cdots\times P(X_2|X_1)\times P(X_1)\\&=P(X_1)\times\prod_{i=2}^{n}P(X_i|X_{i-1},\cdots,X_1)\end{aligned} \tag{11}$$

The construction of a Bayesian networks consists in finding a structure or a graph and estimates its parameters by machine learning. In the case of the classification, the Bayesian networks 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 Figure 1 illustrates its graphical structure.

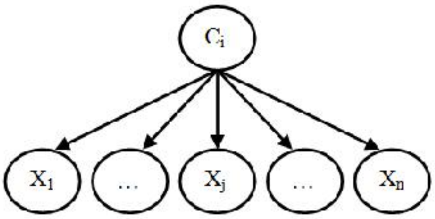


Figure 1. Naive Bayes classifier structure.

To estimate the Bayesian networks 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 [23]:

$$P(X_i|Pa(X_i)) = \frac{1}{\sqrt{2\text{®}\text{±}_i^2}}\exp\left\{\frac{-1}{2\text{±}_i^2}\left(x_i - \left(\text{«}_i + \sum_{j=1}^{n_i}\frac{\text{±}_{ij}}{\text{±}_j^2}(x_j - \text{«}_j)\right)\right)^2\right\} \quad (12)$$

Where,

$Pa(X_i)$ Are the parents of $X_i$;

$\text{«}_i$ , $\text{«}_j$ , $\text{±}_i$ *and* $\text{±}_j$ 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;

$\text{±}_{ij}$ is the regression matrix of weights.

After the parameter and structure learning of a Bayesian networks, 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 Ci is the one that maximizes these probabilities [24, 25]:

$$P(C_i|X) = \begin{cases} P(C_i)\prod_{j=1}^{n} P(X_j|Pa(X_j), C_i) & if \quad X_j \ has \ parents \\ P(C_i)\prod_{j=1}^{n} P(X_j|C_i) & else \end{cases} \quad (13)$$

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 [26]:

$$P(C_i|X) = P(C_i)\prod_{j=1}^{n} P(X_j|C_i) \quad (14)$$

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

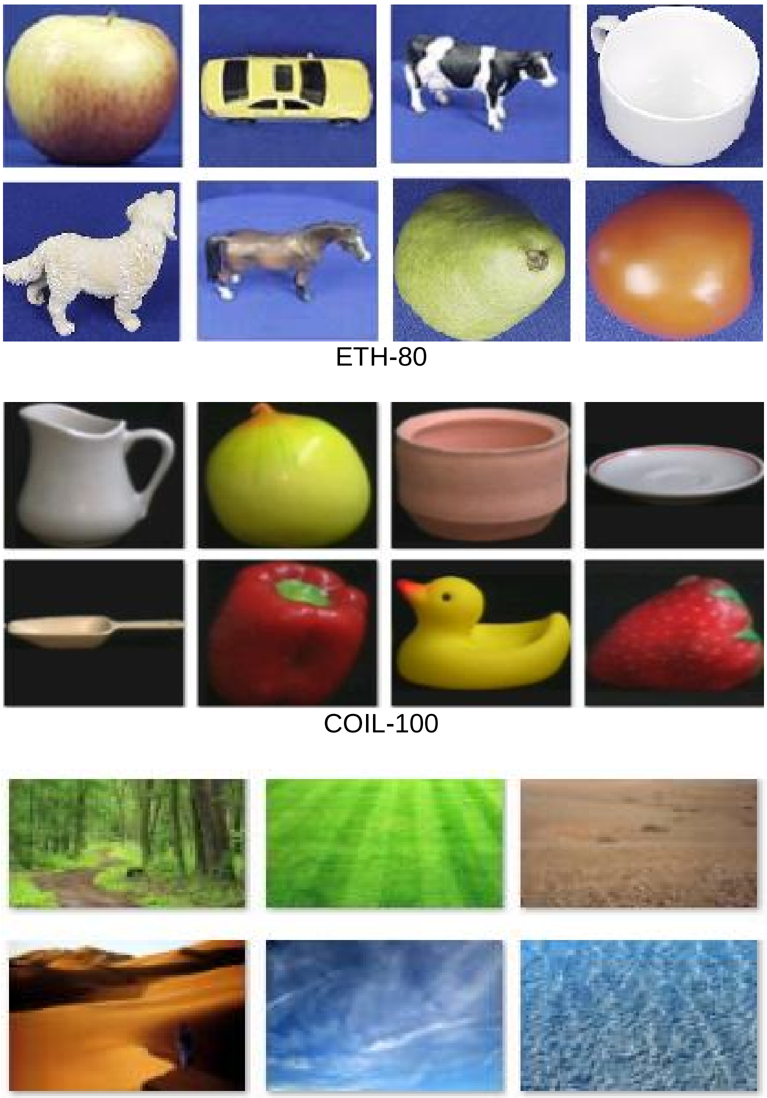
$$\begin{aligned} d(X) &= \arg\max_{C_i} \ P(C_i|X) \\ &= \arg\max_{C_i} \ P(X|C_i)\,P(C_i) \\ &= \arg\max_{C_i} \ P(C_i)\prod_{j=1}^{n} P(X_j|C_i) \end{aligned} \quad (15)$$

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

## 4. Experiments and Results

In In this section we analyze and compare the performance of two approaches for combining low level visual features for semantic image annotation.

In order to achieve this goal, we conduct two experiments on three image databases ETH-80 [14], COL-100 [15] and NATURE [16]. The Figure 2 shows some examples of image objects from these three image databases used in our experiments.



ETH-80

COIL-100

NATURE

Figure 2. Some examples of images for objects from ETH-80, COL-100 and NATURE databases

In the phase of learning and classification, we use a training set of 40 images and a test set of 40 images for each image databases.

In all experiments, the features described in Section 2 are extracted after image segmentation by region growing. For each region that represent an object, 10 Legendre moments (L00, L01, L02, L03, L10, L11, L12, L20, L21, L30) and 16 elements for RGB color histograms are extracted from each color plane namely R,G and B. The number of input features extracted using Texture extraction method is 14 Haralick indices for each one of the 6 co-occurrence matrices hencing 84 textures attributes.

### 4.1. Experiment 1

In this experience, we realize a semantic image annotation using an approach that groups all types of low-level visual features, extracted from a region image, in a single vector used as the input of the Bayesian network classifier.

The experimental approach adopted in this experience is represented by the Figure 3.

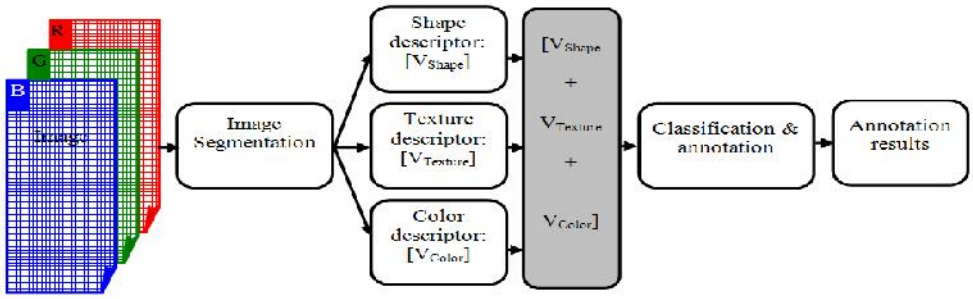

R
G
B
Image
Image Segmentation
Shape descriptor: [V_Shape]
Texture descriptor: [V_Texture]
Color descriptor: [V_Color]
[V_Shape
+
V_Texture
+
V_Color]
Classification & annotation
Annotation results

Figure 3. Experimental approachadopted in experience 1

Table 1 summarizes the global average annotation rate obtained in this experiment for each image database.

Table 1. Global Average Annotation Rate and Error Rate.

| Database | Global average Annotation rate | Error rate |
|---|---|---|
| ETH-80 | 87.50% | 12.50% |
| COIL-100 | 82.50% | 17.50% |
| NATURE | 90.00% | 10% |

Figures 4 and 5 show the confusion matrix.

| | Predicted keyword | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | apple | car | cow | cup | dog | horse | pear | tomato |
| apple | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| car | 0 | 4 | 0 | 1 | 0 | 0 | 0 | 0 |
| cow | 0 | 1 | 4 | 0 | 0 | 0 | 0 | 0 |
| cup | 1 | 0 | 0 | 4 | 0 | 0 | 0 | 0 |
| dog | 0 | 1 | 0 | 0 | 4 | 0 | 0 | 0 |
| horse | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 |
| pear | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 |
| tomato | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |

Figure 4. Confusion matrix for images of database ETH-80.

| | Predicted keyword | | | | | |
|---|---|---|---|---|---|---|
| | forest | gazon | ground | sahara | sky | water |
| forest | 5 | 0 | 0 | 0 | 0 | 0 |
| gazon | 0 | 4 | 1 | 0 | 0 | 0 |
| ground | 0 | 1 | 4 | 0 | 0 | 0 |
| sahara | 1 | 0 | 0 | 4 | 0 | 0 |
| sky | 0 | 0 | 0 | 0 | 5 | 0 |
| water | 0 | 0 | 0 | 0 | 0 | 5 |

Figure 5. Confusion matrix for images of database NATURE

### 4.2. Experiment 2

In this experience, we realize a semantic image annotation using an approach that combines all different types of visual features separately in parallel scheme. Each type of descriptors is used as the input of a Bayesian networks classifier. The annotation decision is realized by the vote of combined classifiers corresponding to the different types of descriptors. The outputs of the classifiers are combined in the decision level to produce a final score for semantic image annotation.

The experimental approach adopted in this experience is represented by the Figure 6.

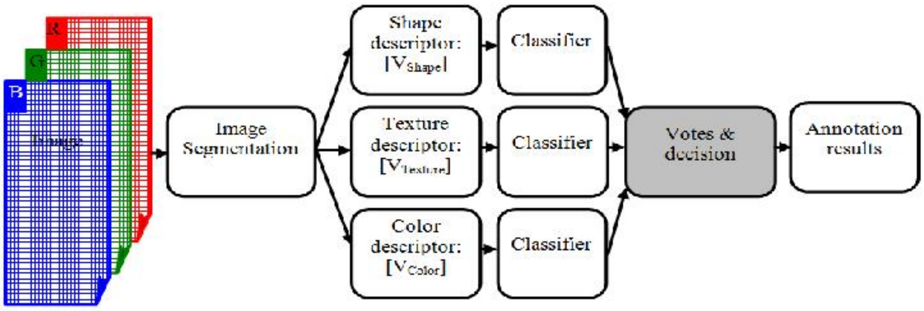

R
G
B
Image
Image Segmentation
Shape descriptor: [$V_{Shape}$]
Texture descriptor: [$V_{Texture}$]
Color descriptor: [$V_{Color}$]
Classifier
Classifier
Classifier
Votes & decision
Annotation results

Figure 6. Experimental approachadopted in experience 2

Table 2 summarizes the global average annotation rate obtained in this experiment for each image database.

Table 2. Global Average Annotation Rate and Error Rate.

| Database | Global average Annotation rate | Error rate |
|---|---|---|
| ETH-80 | 90.00% | 10.00% |
| COIL-100 | 85.00% | 15.00% |
| NATURE | 93.33% | 6.77% |

Figures 7 and 8 show the confusion matrix.

| | Predicted keyword | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | apple | car | cow | cup | dog | horse | pear | tomato |
| apple | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| car | 0 | 4 | 0 | 0 | 0 | 0 | 1 | 0 |
| cow | 0 | 2 | 3 | 0 | 0 | 0 | 0 | 0 |
| cup | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 1 |
| dog | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 |
| horse | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 |
| pear | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 |
| tomato | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |

Figure 7. Confusion matrix for images of database ETH-80

| | Predicted keyword | | | | | |
|---|---|---|---|---|---|---|
| | forest | gazon | ground | sahara | sky | water |
| forest | 5 | 0 | 0 | 0 | 0 | 0 |
| gazon | 0 | 4 | 0 | 0 | 1 | 0 |
| ground | 0 | 0 | 5 | 0 | 0 | 0 |
| sahara | 0 | 0 | 0 | 4 | 1 | 0 |
| sky | 0 | 0 | 0 | 0 | 5 | 0 |
| water | 0 | 0 | 0 | 0 | 0 | 5 |

Figure 8. Confusion matrix for images of database NATURE

### 4.3. Analysis of Results

As can be observed from Tables 1 and 2, the second approach for combining low level visual features including color, texture and shape descriptors, produces the better global average annotation rates for all the tree images databases. Also, analysis of confusion matrix presented by the Figure 4, 5, 7 and 8, shows that the individual annotation rates obtained for some objects (dog, tomato, and ground) with this approach are better than those obtained with the first combining approach. So it appears from these remarks that the combination of different types of low level visual features in parallel scheme will improve the semantic image annotation rates. This rate can be further improved by increasing the number of combined descriptors. In the case of the merger,as done in some work reaserch [1, 2], [8-12], it is not obvious because the size of the vector of the merged descriptors becomes very large and hence selectingappropriate low level features [1, 3, 6], in order to reduce the dimension of combined feature vector, becomes an open problem.

## 5. Conclusion

In this work, we have studied two different approaches for combining low level visual features including color, texture and shape descriptors for semantic image annotation via Bayesian networks classifier.

Analysis and comparative results, obtained from experiments realized on three image databases, have shown that combining the different types of visual features in parallel scheme gives better annotation accuracy than merging these features together in one vector.

Our investigation suggests that the most fruitful approaches, bringing practical benefits for semantic image annotation, will involve an appropriate combination of different types of low level visual features.

In future work, we would like to develop others combination schemes that integrate the joint distribution of multiple features.

## References

[1] Ajimi A, Sreek. Efficient Automatic Image Annotation using Weighted Feature Fusion and its Optimization using Genetic Algorithm. *Communications on Applied Electronics*. 2015; 1(6): 15-19.

[2] Ivan D, Luciano S. *Mixing Low-Level and Semantic Features for Image Interpretation*. Proceedings of Computer Science Computer Vision - Workshops. Zurich, Switzerland. 2015; 8926: 283-298.

[3] Cong J. Automatic image annotation using feature selection based on improving quantum particle swarm optimization. *Signal Processing journal*. 2015; 109: 172-181.

[4] Hengam D, Eskandari A. A Novel semantic statistical model for automatic image annotation Using the Relationship between the Regions Based on multi-criteria Decision Making. *Electrical and computer Engineering*. 2014; 4(1): 37-51.

[5] Zhang J, Da Li, Hu W, Chen Z, Yuan Y. Multilabel Image Annotation Based on Double-Layer PLSA Model. *The Scientific World Journal*. 2014: 1-9.

[6] Dongping Z, Yanjie Li, Huailiang P, Yafei Lu. Image Annotation Based on Joint Feature Selection with Sparsity. *Information Technology Journal*. 2014; 13: 102-109.

[7] Dong P Tian. A Review on Image Feature Extraction and Representation. *Techniques Multimedia and Ubiquitous Engineering journal*. 2013; 8(4): 385-395.

[8] Fernando B, Fromont E, Muselet D, Sebban M. *Discriminative Feature Fusion for Image Classification*. IEEE Conference Computer on Vision and Pattern Recognition (CVPR). 2012: 3434-3441.

[9] Krichna A, Prasad B. Automated image annotation for retrieval of medical images. *Computer Application journal*. 2012; 55 (3): 0975-8887.

[10] Zhang R. Combining visual features and contextual information for image retrieval and annotation. Theses and dissertations. Toronto, Ontario, Canada: Ryerson University; 2011.

[11] Zhang R, Guan l, Zhang L, Xin-Jing W. *Multi-Feature pLSA for Combining Visual Features in Image Annotation*. Proceedings of the 19th ACM international conference on Multimedia. Scottsdale, Arizona, USA. 2011: 1513-1516.

[12] Jian H, Zhang B, Nai-Ming Qi, Yang Y. Evaluating Feature Combination in Object Classification. *Advances in Visual Computing journal*. 2011; 6939: 597-606.

[13] Wang, Mei T, Gong S, Hua X. Combining global, regional and contextual features for automatic image annotation. *Pattern Recognition journal*. 2009; 42(2): 259-266.

[14] ETH-80 database image. Avalable online: http://www.d2.mpi-inf.mpg.de/Datasets/ETH80.

[15] COIL-100 database image. Avalableonline: http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php.
[16] Minaoui B, Oujouara M, Fakir M. Combining Generative And Discriminative Classifiers For Semantic Automatic Image Annotation. *Image Processing journal*. 2014; 8(5): 225-244.
[17] Shih Y, Cheng S. Automatic seeded region growing for color image segmentation. *Image and Vision Computing journal*. 2005; 23: 877-886.
[18] Chonga C, Raveendranb P, Mukundan R. Translation and scale invariants of Legendre moments. *Pattern Recognition journal*. 2004; 37: 119-129.
[19] Haralick R, Shanmugan K, Dinsteinl. Textural features for image classification. *IEEE Transactions on SMC.* 1973; 3(6): 610-621.
[20] Becker A, Naim P. les réseaux bayésiens: modèles graphiques de connaissance. Eyrolles. 1999.
[21] Pearl J. Bayesian Networks. *Handbook of Brain Theory and Neural Networks*.1995: 149-153.
[22] Sabine B. Modèles graphiques probabilistes pour la reconnaissance de formes. Theses. Nancy 2 University; 2009.
[23] George H, Langley P. *Estimating continuous distributions in bayesian classifiers*. The Eleventh Conference on Uncertainty in Artificial Intelligence. 1995.
[24] Leray Ph. Réseaux bayésiens: apprentissage et modélisation de systèmes complexes. These. Rouen University; 2006.
[25] Naïm P, Wuillemin H, Leray P, Pourret O, Becker A. Réseaux bayésiens. 3ème edition. Eyrolles. 2008.
[26] Tom Mitchell. Generative and discriminative classifier: Naïve bayes and logistic regression. *Machine learning*. 2010.