NRC·CNRC
From Discovery
to Innovation...
De la découverte
à l'innovation...

La Science à l'oeuvre pour le Canada
at work for Canada

NRC Publications Archive (NPArC)
Archives des publications du CNRC (NPArC)

**Fuzzy Indexing for Bag of Features Scene Categorization**

Bouachir, Wassim; Kardouchi, Mustapha; Belacel, Nabil

**Web page / page Web**

http://nparc.cisti-icist.nrc-cnrc.gc.ca/npsi/ctrl?action=rtdoc&an=16350480&lang=en
http://nparc.cisti-icist.nrc-cnrc.gc.ca/npsi/ctrl?action=rtdoc&an=16350480&lang=fr

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
http://nparc.cisti-icist.nrc-cnrc.gc.ca/npsi/jsp/nparc_cp.jsp?lang=en
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
http://nparc.cisti-icist.nrc-cnrc.gc.ca/npsi/jsp/nparc_cp.jsp?lang=fr
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Contact us / Contactez nous: nparc.cisti@nrc-cnrc.gc.ca.

National Research
Council Canada

Conseil national
de recherches Canada

Canada

# *Fuzzy Indexing for Bag of Features Scene Categorization*

Wassim Bouachir and Mustapha Kardouchi
Computer Science Department
Moncton University
Moncton, Canada
{wassim.bouachir, mustapha.kardouchi}@umoncton.ca

Nabil Belacel
Institute for Information Technology
National Research Council
Moncton, Canada
nabil.belacel@nrc-cnrc.gc.ca

*Abstract*—**This paper presents a novel Bag of Features (BoF) method for image classification. The BoF approach describes an image as a set of local descriptors using a histogram, where each bin represents the importance of a visual word. This indexing approach has been frequently used for image classification, and we have seen several implementations, but crucial representation choices – such as the weighting schemes – have not been thoroughly studied in the literature. In our work, we propose a Fuzzy model as an alternative to known weighting schemes in order to create more representative image signatures. Furthermore, we use the Fuzzy signatures to train the Gaussian Naïve Bayesian Network and classify images. Experiments with Corel-1000 dataset demonstrate that our method outperforms the known implementations.**

***Keywords-Bag of Features; Image Classification; Fuzzy Assignment; Weighting Schemes; Naïve Bayesian Network.***

## I. Introduction

The expansion of means for acquisition, storage and exchange is producing growing image databases. Managing and accessing such huge collections is becoming a field of great interest for computer vision researchers. In this work, we consider the problem of recognizing the semantic category of an image. For instance, we may want to classify a given image to one of these categories: *Building*, *Mountain*, *beach*, etc. This recognition task requires automatically analyzing, and transforming visual contents into representative features in order to index images.

BoF model is a recent indexing method that uses local descriptors to represent interest regions and consider images as sets of elementary features [1], [2]. The description of an image collection with this approach requires three main steps: detecting and describing interest regions, quantifying extracted local descriptors to build a visual vocabulary, and finally indexing each image by computing a signature that contains the weights of all visual words of the vocabulary. The weights are calculated according to a weighting scheme and each one represents the importance of a visual word in the image. The BoF framework was conceived analogically to the "Bag of Words" approach in text retrieval domain [3], [4], [5]. Consequently, computer vision researchers have been using text retrieval weighting schemes to compute the weights of visual words. Since there are fundamental differences between textual words and visual words, we aim to define a specific weighting scheme for BoF indexing using a Fuzzy model. Our method maintains simplicity and efficiency of the BoF approach, while producing a Fuzzy signature that reflects the real weights of visual words. We also propose to train the Gaussian Naive Bayesian Network using the obtained Fuzzy weights and evaluate our method for scene classification.

The paper is organized as follows: the second section describes the BoF framework, the third one reviews the known weighting schemes and presents shortcomings of such representations. Sections 4 and 5 respectively present the Fuzzy method and the classification model that we propose. Section 6 provides detailed experimental results, and section 7 concludes the paper.

## II. BoF Framework

The BoF model describes each image using a set of visual patterns called visual vocabulary. The vocabulary is obtained by clustering local features extracted from images, where each resulting cluster is a visual word. An image is finally represented by a histogram. Each bin of this histogram corresponds to a visual word, and the associated weight represents its importance in the image. Thereby, the construction of the histogram requires three steps: A) extracting visual features, B) building a visual vocabulary and C) indexing images.

### *A. E xtracting Local Features*

A very interesting approach for extracting local features is to detect keypoints. Those are the centers of salient patches generally located around the corners and edges. In our work, we detect and describe keypoints using Scale Invariant Features Transform (SIFT) [6] because of its reasonable invariance to changes in illumination, image noise, rotation, scaling, and small changes in viewpoint [7]. In this step, SIFT keypoints are extracted, and each one is described by a vector of 128 elements summarizing a local information. The extracted features will be used to build the visual vocabulary.

### *B. Building the Visual Vocabulary*

Building the visual vocabulary means quantifying extracted local descriptors. The vocabulary can be generated by clustering SIFT features using the standard k-means algorithm. The size of the vocabulary is the number of clusters, and the centers of clusters are the visual words. Each image in the database will be represented by visual words from this vocabulary.

978-1-4244-5997-1/10/$26.00 ©2010 IEEE

### C. Image Indexing

Once the visual vocabulary is built, we index each image by constructing its BoF signature. This requires finding the weight of the visual words from the vocabulary. Each image is described by a histogram, where the bins are the visual words and the corresponding values are the weights of the words in the image.

## III. A Review of Weighting Schemes for BoF Indexing

### A. Popular Weighting Schemes

Analogically to the text retrieval approach, the weight of a visual word is obtained by multiplying three factors explained below and detailed in table 1:

- Term Frequency *(tf)*: The visual word is frequently mentioned in an image.
- Inverse Document Frequency *(idf)*: This is a collection-dependent factor used to favour visual words found in a few images and to down-weight those that often appear in the collection.
- The normalization factor: This component is introduced to treat equally all the images, because the number of keypoints varies depending on the complexity of the image content.

For image search, we have seen the use of *term frequency-inverse document frequency* (*tfx)* [1], [8] and the count of visual words *(txx)* [9]. We have also seen the use of *txx* [10] and binary weights *(bxx)* [2] for image classification. Note that all of these methods perform the nearest neighbour search in the vocabulary to map each keypoint to the most similar visual word.

### B. Drawbacks of Existing representations

Using term weighting schemes migrated from text retrieval domain is not the optimal alternative. In fact, the textual terms vocabulary is generated naturally by analyzing the text corpus, while the visual words vocabulary is the output of numerical vector quantization by using the clustering algorithm. Furthermore, a “Bag of Words” vector of a text document is obtained naturally by finding in the vocabulary the word stem in accordance with the language grammar and semantic. A BoF for an image is obtained in a different way by mapping keypoints to visual words. A similarity measure between numerical vectors is used and each keypoint is considered as its nearest visual word from the vocabulary. Indexing images in this way reduces the discriminative power of the signature. Two keypoints may be assigned to the same visual word even if they are not equally similar to this word. Consequently, they contribute in the same way to the construction of the image signature, and the obtained value does not reflect the real weight of the visual word. Certainly, the more the vocabulary size is increased, the more this effect is opposed. But in this case, two similar keypoints may be considered as two different visual words. In addition, the vocabulary would be noise sensitive, less generalizable, and incurs longer processing time to train the classifier. Instead of using a text retrieval weighting scheme, we propose a more realistic approach to weight visual words by using a Fuzzy assignment.

TABLE I
DESCRIPTION OF THE TERM WEIGHTING FACTORS [3]

| Name | Value | Description |
|---|---|---|
| | | ***Term frequency factor*** |
| *b* | 1 or 0 | Binary i.e. 1 for visual words present, 0 if not. |
| *t* | *tf* | Number of occurrence of the visual word. |
| | | ***Collection frequency factor*** |
| *x* | 1.0 | No change in weight. |
| *f* | $log\left(\frac{NC}{nv}\right)$ | Multiply by *idf* (*NC* is the number of images in the collection, and *nv* the number of images containing the visual word). |
| | | ***Normalization factor*** |
| *x* | 1.0 | No normalization. |
| *c* | $\frac{1}{\sum w_i}$ | Each weight $w_i$ is divided by the sum of the image weights. |

## IV. The Fuzzy Representation

Suppose that $V = \{v_1, v_2, ..., v_i, …, v_k\}$ is the vocabulary formed by the *k* centers of clusters (visual words) obtained after vector quantization with k-means algorithm. Let $p_j$, $j \in \{1, 2, …, M\}$ be a SIFT local descriptor among *M* keypoints descriptors extracted from an image. We associate to *pj* a Fuzzy description considering the whole vocabulary. This description represents the contribution of the keypoint in the weight of each visual word. For this purpose, a membership degree is defined using the Fuzzy membership function of Fuzzy-C-Means algorithm [11]:

$$U_{ij} = \frac{1}{\sum_{n=1}^{k}\left(\frac{||p_j - v_i||}{||p_j - v_n||}\right)^{\frac{2}{m-1}}} \qquad (1)$$

where $U_{ij}$ is the contribution of the keypoint described by $p_j$ in the weight of the visual word $v_i$ , and *m* is the degree of fuzziness. Thus, a Fuzzy histogram is obtained and each bin represents the Fuzzy weight of the corresponding visual word. The main advantage of such representation is that it considers the similarity between the keypoint and each visual word from the vocabulary. To illustrate this effect, let us consider two different local descriptors $p_1$ and $p_2$ having the same closest cluster center. In this case, $p_1$ and $p_2$ contribute in the same way to the weight of their nearest visual word even if they are not equally similar to this word. By using the Fuzzy assignment, the two keypoints contribute to the weights of all the similar words, and thus the distribution is more equitable. The parameter *m* $(1<m<\infty)$ controls the degree of fuzziness in the distribution of weights. Empirically, we found that *m*=1.1 is the best setting.

## V. Categorization by Naïve Bayes

The Naïve Bayesian Network (NBN) has been widely used for bags of words text categorization because of its simplicity, learning speed and competitiveness with the state-of-the art classifiers [4], [5], [13], [12]. Consequently, it has also been used as a BoF image classifier [10]. The main idea of this model is to learn from a training set the conditional probability of each attribute given a class. The classification decision is taken by applying Bayes’s rule:

$$P(C_i|X_n) = \frac{P(C_i)\,P(X_n|C_i)}{P(X_n)} \qquad (2)$$

where $P(C_i|X_n)$ is the probability of the category $C_i$ given $X_n$ (the BoF vector of an image $I_n$). $P(C_i)$ and $P(X_n)$ are respectively the prior probability of the class $C_i$, and the prior probability of obtaining the signature $X_n$ for an image. The probability $P(X_n)$ is the same for all the classes, and therefore, it can be ignored without affecting the relative values of class probabilities. Finally, we consider the largest a posteriori score as the class prediction. This prediction is possible by making a strong independence assumption called the *naïve assumption*: the visual words of the vocabulary are conditionally independent given the class. The reason why NBN is able to work well with the BoF approach is that the conditional independence assumption is quite reasonable: if we know that an image belongs to a category, this is sufficient to specify what kind of visual words we will find in this image. Moreover, BoF approach uses high-dimensional attribute spaces where it is very difficult to estimate the correlation between attributes. Practically, attributes are seldom independent given the class, but it has been verified that the NBN performs well even when strong attribute dependences are present [14]. The other important aspect that motivated our classifier choice is its tolerance to learn parameters from different data types generated by different weighting schemes. In existing works, we have seen the use of *txx* [10] and binary weights *(bxx)* [2] for image classification. To compare the weighting schemes performance, we train two instances of NBN. The first learns its parameters from data produced by applying *bxx*, while the second uses *txx* data. Further, we use the Gaussian NBN to learn from the Fuzzy weights.

### A. Conditional Probabilities Estimation for Binary Weights

With *bxx*, the BoF vector of an image $I_n$ is $X_n$=($w_1$,..., $w_j$..., $w_k$) where $w_j$ is the weight of $x_j$ (the *j*th visual word in the vocabulary). The weight $w_j$ is 1 if the word is present, and 0 if not. Given the *naïve assumption* explained above, the conditional probabilities for these binary attributes are computed from the frequencies by counting the number of occurrences of each possible attribute value with each class. Categorization is done by applying equation (2) after decomposing $P(X_n|C_i)$ into the product of the conditional probabilities learned for each attribute value:

$$P(C_i|X_n) = P(C_i) \prod_{j=1}^{k} P(w_j = v|C_i) \qquad (3)$$

with $v \in \{0, 1\}$. Note that in order to avoid probabilities of zero, $P(w_j = v|C_i)$ are computed with Laplace smoothing:

$$P(w_j = v|C_i) = \frac{(\#\ images\ of\ class\ C_i\ with\ w_j = v) + 1}{(\#\ images\ of\ class\ C_i) + 2} \qquad (4)$$

### B. Multinomial Naïve Bayes for txx Representation

The multinomial NBN has been widely used for text classification, where a document is represented by the set of stems occurrences [4], [5], [12], [15]. With *txx* features, the BoF vector contains the visual words counts so that we can model the classifier parameters using the multinomial distribution. During learning step, the classifier computes the relative visual words probabilities separately for each class as follows:

$$P(x_j|C_i) = \frac{N_{ij} + 1}{N_i + k} \qquad (5)$$

where $N_{ij}$ is the count of the visual word $x_j$ in all the training images belonging to class $C_i$, and $N_i$ the count of all visual words in the training images belonging to $C_i$ . Laplace estimator is used as well as in Equation (4) to avoid the zero probability problem. To categorize a new image $I_n$, the Naïve Bayes defines a multinomial distribution by using the vector of $k$ probabilities $P(x_j|C_i)$ for the corresponding class, and by using $N_n$, the number of visual words for that image. The classification is based on the relative frequencies $w_{jn}$ of the visual words in $I_n$ , by multiplying the class prior $P(C_i)$ by $P(X_n|C_i)$. The latter parameter is the probability of obtaining the signature $X_n$ for an image belonging to $C_i$ . This is calculated by using the multinomial mass function, and thus, we get the a posteriori classes score:

$$P(C_i|X_n) = P(C_i)\, N_n! \prod_{j=1}^{k} \frac{P(x_j|C_i)^{w_{jn}}}{w_{jn}!} \qquad (6)$$

Note that we can delete the computationally expensive terms $N_n!$ and $w_{jn}!$ without any change in the results since neither depends on class $C_i$.

### C. The Proposed Gaussian NBN

By using the Fuzzy weighting scheme, we obtain a BoF vector of real valued attributes that represent the Fuzzy weights of visual words. To model the conditional probabilities distributions, we assume that for a given class $C_i$, the Fuzzy weight of each visual word $x_j$ is a normally distributed random variable with mean $\mu_{ij}$ and variance $\sigma_{ij}{}^2$. This model is based on the assumption that for the images belonging to same class, the weights of a visual word tend to cluster around the mean value. The a posteriori score of classes is then computed using Equation (3) with:

$$P(w_j = v|C_i) = \frac{1}{\sqrt{2\pi}\,\sigma_{ij}}\, e^{-\frac{(v-\mu_{ij})^2}{2\sigma_{ij}{}^2}} \qquad (7)$$

where $v \in [\mathbf{0}\ ;\ \infty[$.

## VI. Experiments

We explored the performance of the proposed method on the NBN categorization task conducted on Corel-1000 database[1]. Corel is a collection of about 60000 images created by the professor Wang's group at Penn State University. Corel-1000 is a well known sub-collection that contains 1000 natural images divided into 10 categories with 100 images per category. We extracted SIFT keypoints from Coil-1000 database and we used the k-means clustering algorithm to cluster the extracted local features into a visual vocabulary. For our experiments, we set the size of the vocabulary to 100 visual words. Since previous works relied on binary weights *(bxx)* [2] and term frequency *(txx)* [10] for image classification, we applied these two schemes and the Fuzzy method to index the image collection in three ways. We divided the collection at random into two sets: 700 images are used for training each of the three NBN instances, and 300 images are used for testing. The table 2 shows that when the Gaussian NBN was used with the Fuzzy weighting scheme, 60% of the images were correctly classified, and this was the best rate. With the multinomial NBN, 57% of the scenes were

[1] Available at: http://wang.ist.psu.edu/docs/related.shtml

correctly recognized, whereas the binary weights classifier had the worst percentage (37%).

Corel-1000 is a very challenging collection because of the large number of classes and the high variability of poses and background even for images belonging to the same class. Nevertheless, the conducted experiments demonstrated that when the Gaussian NBN learns from Fuzzy weights, we can handle better difficult situations such as multiple objects in the scene and variable orientation as we can see in figure 1. This figure presents examples where scenes were well classified.

The confusion matrix of the Gaussian NBN is given in table 3 where the diagonal elements show interesting correct classification rates for most of classes. It also shows two high rates obtained for the classes *Dinosaur* and *Flowers* (respectively 92% and 94%). The lowest rates are 41% and 40%, and were obtained respectively for the categories *Building* and *Mountain*. The last two percentages could be explained by the fact that these two categories are sharing objects with other classes. For example, 17% of *building* scenes were confused with the category *Bus* because many images from the latter contain also buildings.

## VII. Conclusion

We presented an efficient implementation of the BoF image classification approach, and we demonstrated that the classical text representation techniques are not a suitable choice for images. The proposed method relies on a Fuzzy model for visual indexing and uses the Gaussian NBN for image classification. The BoF indexing could be improved by several other ways, such as using a more effective algorithm to create the visual vocabulary. In fact, a more representative vocabulary would be generated by using a clustering algorithm that handles the large number of local descriptors and the presence of outliers. On the other hand, SIFT descriptors use only gray scale information, while the color provides valuable information in keypoints description. This proposes a further improvement by introducing the color information to describe keypoints. One other interesting direction for future work would be to divide the image signature into sub-histograms. Each sub-histogram would correspond to a part of the described image. As a result, the BoF signature is enriched by the information on the spatial relation among visual words.

TABLE II
Classification Rates

| Weighting scheme | Percentage of correct classification |
|---|---|
| *bxx* | 37% |
| *txx* | 57% |
| *Fuzzy weights* | 60% |

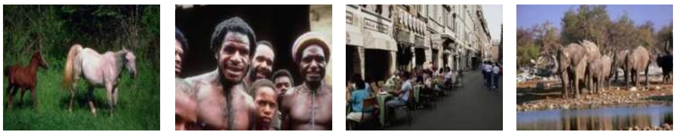

Fig. 1. Scenes correctly classified by the Gaussian NBN as: *Horse*, *Africa*, *Building*, *Elephant*.

TABLE III
Confusion Matrix (in Percentages) for the Gaussian NBN

| ↓ True classes | *Africa* | *Beach* | *Building* | *Bus* | *Dinosaur* | *Elephant* | *Flowers* | *Horse* | *Mountain* | *Food* |
|---|---|---|---|---|---|---|---|---|---|---|
| *Africa* | **46** | 0 | 5 | 8 | 0 | 3 | 14 | 3 | 11 | 11 |
| *Beach* | 0 | **45** | 10 | 0 | 0 | 7 | 17 | 3 | 7 | 10 |
| *Building* | 10 | 10 | **41** | 17 | 0 | 7 | 3 | 0 | 7 | 3 |
| *Bus* | 4 | 0 | 4 | **81** | 0 | 0 | 4 | 0 | 4 | 4 |
| *Dinosaur* | 0 | 8 | 0 | 0 | **92** | 0 | 0 | 0 | 0 | 0 |
| *Elephant* | 0 | 16 | 6 | 0 | 3 | **55** | 3 | 13 | 3 | 0 |
| *Flowers* | 0 | 0 | 0 | 0 | 3 | 0 | **94** | 0 | 3 | 0 |
| *Horse* | 3 | 0 | 3 | 7 | 0 | 0 | 3 | **67** | 13 | 3 |
| *Mountain* | 8 | 4 | 0 | 8 | 4 | 4 | 12 | 0 | **40** | 20 |
| *Food* | 15 | 0 | 15 | 9 | 0 | 6 | 0 | 0 | 9 | **47** |

## References

[1] J. Sivic and A. Zisserman, Video Google: “A Text Retrieval Approach to Object Matching in Videos,” Proceedings of the Ninth IEEE International Conference on Computer Vision: vol.2, 1470-1477, 2003.

[2] E. Nowak, F. Jurie and B. Triggs, “Sampling strategies for bag-of-features image classification,” Proceedings of the European Conference on Computer Vision: 490-503, 2006.

[3] G. Salton and C. Buckley. “Term-weighting approaches in automatic text retrieval,” Information Processing and Management: an Int'l Journal: 24(5), 513-523, 1988.

[4] A. Juan and H. Ney, “Reversing and Smoothing the Multinomial Naive Bayes Text Classifier,” Proceedings of the 2nd Int. Workshop on Pattern Recognition in Information Systems: 200-212, 2002.

[5] A. McCallum, K. Nigam, “A Comparison of Event Models for Naive Bayes Text Classification,” Proceedings of the AAAI-98 Workshop on Learning for Text Categorization: 41-48, 1998.

[6] David G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” The International Journal of Computer Vision: 91-110, 2004.

[7] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” IEEE Transactions on Pattern Analysis & Machine Intelligence, Volume 27, Number 10: 1615-1630, 2005.

[8] W. Zhao, Y. Jiang and C. Ngo, “Keyframe retrieval by keypoints: Can point-to-point matching help?” Proceedings of the 5th international Conference on Image and Video Retrieval: 72-81, 2006.

[9] S. Newsam and Y. Yang, “Comparing Global and Interest Point Descriptors for Similarity Retrieval in Remote Sensed Imagery,” Proceedings of the 15th International Symposium on Advances in Geographic Information Systems, 2007.

[10] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints,” In Workshop on Statistical Learning in Computer Vision, ECCV: 1-22, 2004.

[11] J. Bezdeck, Pattern recognition with fuzzy objective function algorithms, Plenum Press Ed., New-York, 1981.

[12] E. van Dyk and E. Barnard, “Naive Bayesian classifiers for multinomial features: a theoretical analysis,” Proceedings of the Eighteenth Annual Symposium of the Pattern Recognition Association of South Africa: 87-92, 2007.

[13] David D. Lewis, “Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval,” Proceedings of the 10th European Conference on Machine Learning: 4-15, 1998.

[14] P. Domingos and M. Pazzani, “On the Optimality of the Simple Bayesian Classifier under Zero-One Loss,” Machine Learning 29: 103-130, 1997.

[15] S.-B. Kim, H.-C. Rim and H.-S. Lim, “A new method of parameter estimation for multinomial naive bayes text classifiers,” Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval: 391-392, 2002.