FEATURE VECTOR APPROXIMATION BASED ON WAVELET NETWORK

Mouna Dammak, Mahmoud Mejdoub, Mourad Zaied and Chokri Ben Amar
Regim Research Group on Intelligent Machines, ENIS University, BP W, 3038 Sfax, Tunisia

Keywords: Wavelet network, Approximation, Local feature, Bag of words.

Abstract: Image classification is an important task in computer vision. In this paper, we propose a new image representation based on local feature vectors approximation by the wavelet networks. To extract an approximation of the feature vectors space, a Wavelet Network algorithm based on fast Wavelet is suggested. Then, the K-nearest neighbor (K-NN) classification algorithm is applied on the approximated feature vectors. The approximation of the feature space ameliorates the feature vector classification accuracy.

1 INTRODUCTION

Visual descriptors for image categorization generally consist of either global or local features. The former ones represent global information of images. On the contrary, local descriptors (Piro et al., 2010; Tao et al., 2010; Li and Allinson, 2008; Mejdoub et al., 2008; Mejdoub et al., 2009; Mejdoub and BenAmar, 2011) extract information corresponding to locations of a specific image that are relevant to characterize the visual content. Indeed, these techniques are able to emphasize local patterns, which images of the same category are expected to share. Research suggests that a regular dense sampling of descriptors can provide a better representation (Lazebnik et al., 2006; Nowak et al., 2006) than the "interest" points. The bag of words (Csurka et al., 2004) representation can be considered as the practical proof of the effectiveness of visual feature points. This approach is applied to images in (Csurka et al., 2004; Lazebnik et al., 2006) to extract a histogram of words from the image. The essential characteristic of this representation is that it dismisses any kind of information associated to the arrangement of words.

Analyzing a signal from its corresponding graph does not give us access to all the information it contains. It is often necessary to transform it, i.e., to give it another representation which clearly shows its features. Fourier (Fourier, 1822), suggests that all functions must be able to express themselves in a simple way as a sum of sinus. As an advanced alternative to the classical Fourier analysis, wavelets (Cohen et al., 2001) have been successfully used for signal approximation. The fundamental idea behind wavelets is to process data at different scales or resolutions. In such a way, wavelets provide a time-scale presentation of a sequence of input signal (Yan and Gao, 2009). The wavelet transform applies a multi-resolution analysis to decompose a signal into the low-frequency coefficients and the high-frequency coefficients. The former represent the original signal approximation and the latter represent the detailed information of the original signal.

Besides, the Wavelet Networks (WN) (Li et al., 2003) is a powerful tool to approximate signals. Indeed, it mixes the performances of the wavelet theory in terms of localization and multi-resolution representation and the Neural Network in terms of classification. The keypoint of the wavelet networks lies in the optimization of network weights that permits to extract an approximation of the original signal. In (Jemai et al., 2011), the authors introduce a new training method for WN to assess this algorithm in the field of images classification directly from pixel value images. In (Ejbali et al., 2010), the authors advance a new approach for the approximation of acoustic units for the task of the speech recognition.

In this paper, we propose a novel image categorization approach based on the approximation of local features by Wavelet Networks. Firstly, we extract local features based on SIFT (Lowe, 1999) and SURF (Bay et al., 2006) descriptors and we represent them using the BOW model based on spatial pyramid. Secondly, we approximate the histogram of words by wavelet networks in an attempt to obtain greater representation efficiency of the histogram of words. Fi-

Dammak M., Mejdoub M., Zaied M. and Ben Amar C..
FEATURE VECTOR APPROXIMATION BASED ON WAVELET NETWORK.
DOI: 10.5220/0003776803940399
In *Proceedings of the 4th International Conference on Agents and Artificial Intelligence* (ICAART-2012), pages 394-399
ISBN: 978-989-8425-95-9

nally, k-NN algorithm is applied on the approximated histogram of words for image categorization.

The rest of the paper is organized as follows: Section 2 focuses on the theoretical concept of the wavelet networks. Section 3 outlines the proposed approach of the extraction and the approximation of the local descriptors. Some experimental results are presented in the final section with the aim to illustrate the effectiveness of the proposed categorization method.

2 WAVELET NETWORK

The concept of wavelet networks was proposed firstly by Zhang and Benveniste (Zhang and Benveniste, 1992). The basic idea of wavelet networks is to combine the localization property of wavelet decomposition and the optimization property of neural networks learning. The multilayered networks allow the representation of a nonlinear function by training while comparing their inputs and their outputs. This training is made while representing a nonlinear function by a combination of activation functions. The admissible wavelet is used as an activation one. They reached the result that the wavelet networks preserve the property of universal approximation of the RBF networks.

Wavelet analysis gives a representation of signals that simultaneously shows the location in time and frequency, thereby facilitating the physical characteristics identification of the signal source (Morlet et al., 1982). This analysis uses a family of translate-dilated functions $\psi_{a,b}$ constructed from a function ψ of $L^2(\Re)$, called mother wavelet $\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)$ with a, b represent respectively the dilation and translation parameters. Discret Wavelet Transform DWT is defined as a set of wavelets are generated by considering only a sampled value of a and b parameters. For analyzing a signal containing a_0^j points ($1 \leq j \leq m$, j represents the scale parameter) we use only the family wavelets: $\psi\left(a_0^{-m}x - nb_0\right)$ with $n = 1\ldots a_0^{m-j}$:

$$w_{j,n} = a_0^{-\frac{j}{2}} \sum_x f(x)\,\psi\left(a_0^{-j}x - nb_0\right) \qquad (1)$$

For the particular case, when $a_0 = 2$ and $b_0 = 1$, the sampling is called dyadic.

The multiresolution analysis consists on, firstly, a scaling function $\phi(x) \in L^2(\Re)$, which constitutes an orthonormal basis by varying its position on a given scale j. The functions of every scale generate an approximation of a given signal f to analyze. Secondly, additional functions, i.e. wavelet functions, are then used to encode the difference in information between adjacent approximations (Meyer, 1990). If we have a finished number N_w of wavelets $\psi_{a,b}$ obtained from the mother wavelet and a finished number N_s of scales $\phi_{a,b}$ obtained from the mother scaling function ϕ, Eq. 2 will be considered as an approximation of the inverse transform:

$$f(x) \simeq \sum_{j=1}^{N_w} \alpha^j \psi_j(x) + \sum_{k=1}^{N_s} \beta^k \phi_k(x) \qquad (2)$$

This equation establishes the idea of Wavelet Networks.

The model, introduced by Zhang et Benveniste (Zhang and Benveniste, 1992), is composed of three layers. Input of this is considered a set of parameters t_j that describe signal coordinate positions to analysis. So the entries are not actual data but only values describing specific positions of the analyzed signal. The hidden layer contains a set of neurons, each neuron composed of a translated and dilated wavelet. The output layer contains one neuron which sums the outputs of the hidden layer by weighted connections weights $a_j$and d_j that represent respectively the wavelets and the scaling functions coefficients 1. Figure 1 displays the structure of wavelet network of the second model.

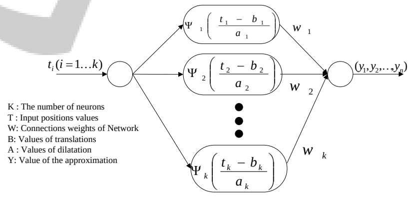


Figure 1: The model of Wavelet Network architecture.

3 OVERVIEW OF THE PROPOSED APPROACH

We suggest in this paper a solution of image classification based on local feature and approximation vector by wavelet networks. The solution which we present proceeds, at first, by local feature image representation. Second, an image is represented based on a BoW model. Third, the approximation by wavelet networks is used on this image. In classification stage (on-line), the same procedure is carried out to extract the approximate test image signature. Finally, to decide upon the image category, we search for the test image the k-similar images in the training data and we apply the majority vote. This search is based on the computation of the distances between the approximate feature vector of the test image and all the approximate feature vectors of the training images. The pipeline of all these stages is illustrated in figure 2.

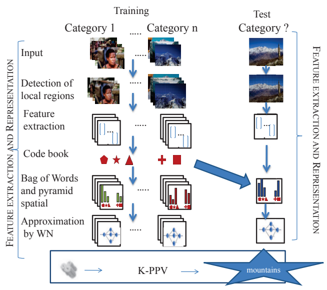


Figure 2: Complete Framework of image classification by BOW and WN.

3.1 Extraction of Local Features

In this paper, we focus on local feature vectors extraction. Our system makes use of three types of low-level features: SIFT features (Lowe, 1999), SIFT-HSV (Bosch et al., 2008) features and SURF (Bay et al., 2006) features. We use a dense sampling to extract patches at a regular grid in the image, and at multiple scales. Given the feature space, the visual vocabulary is built through the clustering of low-level feature vectors using k-means based on the accelerated ELKAN (Elkan, 2003) algorithm for optimization. The clusters define the visual vocabulary and then the image is characterized with the number of occurrences of each visual word. Similar to (Lazebnik et al., 2006) we use a spatial pyramid of $1x1$, $2x2$, and $4x4$ regions in our experiments for all visual features.

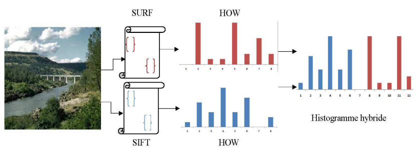


Figure 3: Hybrid descriptor.

The steps needed to calculate the signature of an image for a given local descriptor ($D \in \{SIFT, SURF\}$) are :

1. Extraction of local descriptors based on the descriptor D.
2. Translation of each local descriptor in a histogram of visual words using the technique of bag of words on the descriptors obtained in step (1).
3. Division of the image into bands using the spatial pyramid technique.
4. For each band obtained from the spatial pyramid. extraction of a visual histogram of keywords by combining visual keyword histograms obtained in step (2).
5. Combining histograms obtained for each band to derive a histogram H_D associated with local descriptor D.
6. Combining histograms obtained for SURF descriptor and SIFT descriptor.

3.2 Wavelet Network

The Wavelet Network is used to approximate each local feature vectors computed as indicated in the section 3.1. We denote by V the local feature vector extracted from a given image database. To define the Wavelet Network, we first take a family of n wavelets $\Psi = (\psi_1, \ldots, \psi_n)$ with different parameters of scaling and translation (generated by distributing the parameters on a dyadic grid) that can be chosen arbitrarily at this point. The architecture of the wavelet network is exactly specified by the number of particular wavelets required. In this work, we build the candidate hidden neurons representing a library of wavelets (scaling functions), and then we select the hidden neurons in order to form the optimal structures.

3.2.1 Wavelet Network Initialization

The G library of wavelet and scaling function candidates to join the network are the results of a sampling on a dyadic grid of the parameters of expansion and translation. This family of functions is:

$$\left\{ \begin{array}{c} G = \left\{ \psi\left(a_0^{-m}x - nb_0\right), \phi\left(a_0^{-j}x - nb_0\right) \right\} \\ \text{with } m \in S_a,\ n \in S_{b(m)} \end{array} \right\} \quad (3)$$

In Eq.3, a_0, $b_0 > 0$ are two scalar constants defining the discretization step sizes for dilation and translation. a_0 is typically dyadic. S_b and S_a are finite sets related to the size of the data input domain D. The first derivate from the wavelet beta (Amar et al., 2005) is used as mother wavelet, given by equation 4:

$$\left\{ \begin{array}{cc} B(x) = \left(\frac{x - x_0}{x_c - x_0}\right)^p \left(\frac{x - x_0}{x_c - x_0}\right)^q & if\ x \in [x_0, x_1] \\ 0 & otherwise \\ where p, q, x0, x1 \in \mathbb{R} & and\ x0 < x1 \\ and & x_c = \frac{px_1 + qx_0}{p + q} \end{array} \right\} \quad (4)$$

We give below the used steps to initialize the WN:
Step 1: Start the learning by preparing a library of candidate wavelets and scaling functions.
Step 2: Calculation of the weights corresponding to all the functions of activation of the library.

Step 3: Impose a stop criterion; an error E between the feature vector V and the exit of the network.
Step 4: Initialize the output network to $\tilde{V} = 0$.

3.2.2 Learning Wavelet Network

The training model is built according to the following steps:
Step 1: Calculate the weights γ^l $(\alpha^l$ ou $\beta^l)$ respectively associated with the wavelets and scales functions, of the library already created in section 3.2.1.
Step 2: Calculate the contribution of all the functions of activation in the library $(\gamma^l g_l)$ for the rebuilding of the feature vector V. V can be written :

$$V = \sum_{l=1}^{L} \gamma^l g_l \quad (5)$$

Step 3: Select the function g_k of activation which provided the best approximation of the feature vector V.
Step 4: Recruit the best function with the hidden layer of the wavelet network. The total approximation, which is the accumulation of all the approximations obtained with each iteration, is calculated by: $\tilde{V} = \tilde{V} + \gamma^k g_k$.
Step 5: Compute the difference between the original signal and that approximated one is calculated. If the error E, is reached, then it is the end of the phase of training. If not, we move back to Step 3.

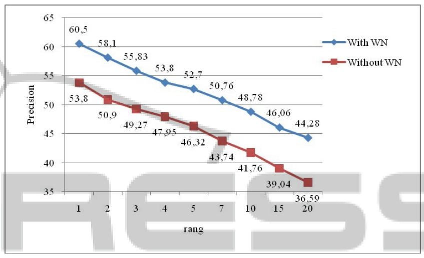
4 EXPERIMENTAL RESULTS

4.1 Presentation of the Datasets

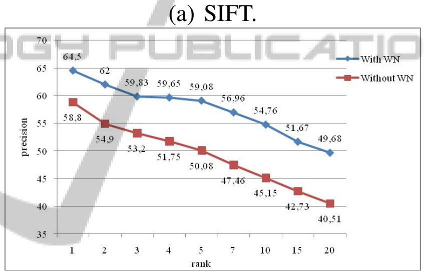
We have carried out experiments on the WANG database (Wang et al., 2001) which contains 1000 images. The database contains ten clusters representing semantic generalized meaningful categories.Besides, we have carried our experiments on the OT dataset (Oliva and Torralba, 2001) that contains 2688 color images , divided in 8 categories. For evaluation we divide the images randomly into 500 training and 500 test images for wang dataset and 800 training and 1688 test images for OT. Experiments were performed on a personal computer with configurations: Intel Core2 Duo (2 GHZ), 4GO. In our experiments, image patches are extracted on regular grids at 4 different scales. SIFT, SIFT-HSV and SURF descriptors are computed at every point of the regular grid. The dense features are vector quantized by the bag of words model with N-cluster = 120 for Wang database and N-cluster = 300 for OT database. We tested the performance of our proposed image retrieval method taking into account the retrieval process accuracy. For the accuracy evaluation, we use the precision.

4.2 Impact of the Space Approximation Feature

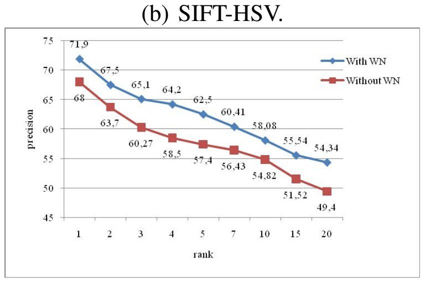
Several experiments are made using the Wang and OT databases. We observe in Figure 5 and 6 that the image signature approximated by WN for the various types of descriptors (SIFT, SIFT-HSV, SURF and Hybrid) give a better performance in term of precision of retrieval than the original signature (signature not approximated by WN).



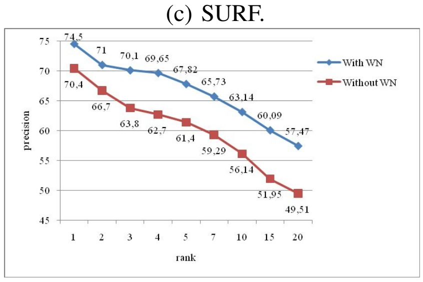
(a) SIFT.



(b) SIFT-HSV.

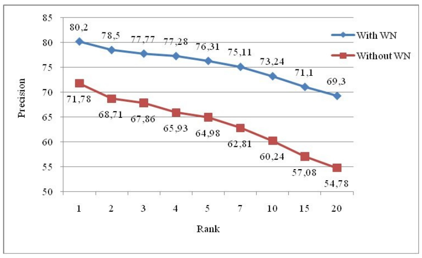


(c) SURF.

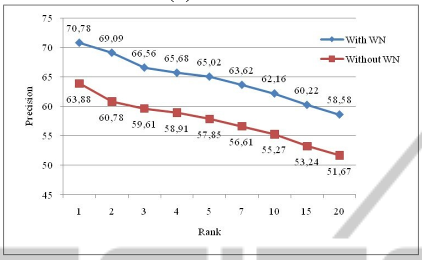


(d) Hybrid.

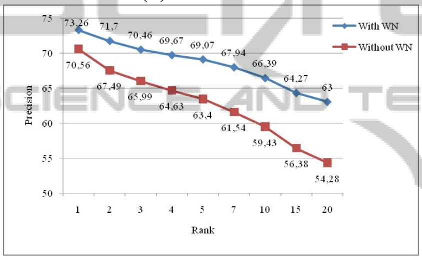
Figure 4: Comparison between original and approximate feature vectors (Wang).



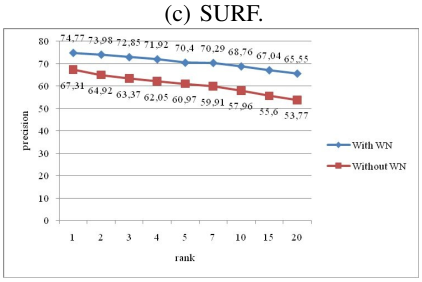
(a) SIFT.



(b) SIFT-HSV.



(c) SURF.



(d) Hybrid.

Figure 5: Comparison between original and approximate feature vectors (OT).

4.3 Comparison of the Proposed Classifier to Popular Methods in the Literature

We will now compare the results and using different machine learning algorithms as Support Vector Machines (SVMs), Hidden Markov Model (HMM), K-Nearest Neighbor (KNN), or Universal Nearest Neighbors rule (UNN) could be applied for categorization.

The experimental results reported here see (1) seem very promising and the proposed approach outperforms the other methods.

Table 1: Comparison with some State-of-the-Art methods.

(a) Wang database.

Classification Model	Classification rate
(Jemai et al., 2011)	71,2
(Jemai et al., 2010)	71,4
(Mouret et al., 2009)	70,60
Our approach	74,50

(b) OT database.

Classification Model	Classification rate
(Piro et al., 2010)-KNN	73,8
(Piro et al., 2010)-UNN	75,70
(Oliva and Torralba, 2001)	83,70
(Horster et al., 2008)	79
Our approach	80,20

4.4 Comparison between Raw Pixels and Feature Vectors Representation

In (Jemai et al., 2011), the authors reshape the size of the image to 90*90 pixels. The wavelets network is directly applied to the pixel values of image.Our work allows on the one hand reducing the dimension of the feature vector of the image, and on the other hand generating a compact representation of the image. And consequently, we obtained a faster computing time and a more precise categorization rate. Table 2 compares CPU-times spent to compute the different stages of the learning algorithm for one image and rate categorization.

Table 2: Classification rate and Time consumption for the processing steps on Wang.

	BWNN	(Jemai et al., 2011)	Our approach
Time consumption			
Training	20mn	0.1176s	0,010531s
Classification	2mn	0.0627s	0.038517s
Classification rate			
Rate classification	60,2	71,4	74,5

5 CONCLUSIONS

A new indexing method was proposed in this paper. This method is based on a combined local feature extraction and approximated signal by Wavelet Network is proposed. This method is applied to the image classification fields. Based on the experiment results, the proposed approach exhibits high classification rates and small computing times.

REFERENCES

Amar, C. B., Zaied, M., and Alimi, M. A. (2005). Beta wavelets. synthesis and application to lossy image compression. *Advances in Engineering Software*, 36:459–474.

Bay, H., Tuytelaars, T., and Gool, L. V. (2006). Surf: Speeded up robust features. In *9th European Conference on Computer Vision*.

Bosch, A., Zisserman, A., and Munoz, X. (2008). Scene classification using a hybrid generative/discriminative approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(4):712–727.

Cohen, A., Dahmen, W., Daubechies, I., and Devore, R. (2001). Tree approximation and optimal encoding. *Applied and Computational Harmonic Analysis*, 11(2):192–226.

Csurka, G., Dance, C. R., Fan, L., Willamowski, J., and Bray, C. (2004). Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*.

Ejbali, R., Zaied, M., and Amar, C. B. (2010). Wavelet network for recognition system of arabic word. *International Journal of Speech Technology*, 13(3):163–174.

Elkan, C. (2003). Using the triangle inequality to accelerate k-means. *ICML*, pages 147–153.

Fourier, J. B. J. (1822). *Thorie analytique de la chaleur*.

Horster, E., Greif, T., Lienhart, R., and Slaney, M. (2008). Comparing local feature descriptors in plsa-based image models. In *DAGM-Symposium*, pages 446–455.

Jemai, O., Zaied, M., Amar, C. B., and Alimi, M. A. (2010). Fbwn: An architecture of fast beta wavelet networks for image classification. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, Barcelona.

Jemai, O., Zaied, M., Amar, C. B., and Alimi, M. A. (2011). Pyramidal hybrid approach: Wavelet network with ols algorithm-based image classification. *International Journal of Wavelets, Multiresolution and Information Processing (IJWMIP)*, 9(1):111–130.

Lazebnik, S., Schmid, C., and Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2169–2178, New York.

Li, C., Liao, X., and Yu, J. (2003). Complex-valued wavelet network. *Journal of Computer and System Sciences*, 67(3):623–632.

Li, J. and Allinson, N. M. (2008). A comprehensive review of current local features for computer vision. *Neurocomputing*, 71:1771–1787.

Lowe, D. (1999). Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, pages 1150–1157, , Corfu, Greece.

Mejdoub, M. and BenAmar, C. (2011). Hierarchical categorization tree based on a combined unsupervised-supervised classification. In *Seventh International Conference on Innovations in Information Technology*.

Mejdoub, M., Fonteles, L., BenAmar, C., and Antonini, M. (2008). Fast indexing method for image retrieval using tree-structured lattices. In *Content based multimedia indexing CBMI*.

Mejdoub, M., Fonteles, L., BenAmar, C., and Antonini, M. (2009). Embedded lattices tree: An efficient indexing scheme for content based retrieval on image databases. *Journal of Visual Communication and Image Representation, Elsevier*.

Meyer, Y. (1990). *Ondelettes et oprateurs I*. Actualits Mathmatiques Current Mathematical Topics. Hermann, Paris.

Morlet, J., Arens, G., Fourgeau, E., and Giard, D. (1982). Wave propagation and sampling theory. *Geophysics*, 47:203–236.

Mouret, M., Solnon, C., and Wolf, C. (2009). Classification of images based on hidden markov models. In *IEEE Workshop on Content Based Multimedia Indexing*, pages 169–174.

Nowak, E., Jurie, F., and Triggs, B. (2006). Sampling strategies for bag-of-features image classification. In *European Conference on Computer Vision*. Springer.

Oliva, A. and Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175.

Piro, P., R.Nock, Nielsen, F., and Barlaud, M. (2010). Boosting k-nn for categorization of natural scenes. *CoRR*.

Tao, Y., Skubic, M., Han, T., Xia, Y., and Chi, X. (2010). Performance evaluation of sift-based descriptors for object recognition. In *Prooceeding of The International MultiConference of Engineers and Computer Scientists*.

Wang, J. Z., Li, J., and Wiederhold, G. (2001). Simplicity : Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):947–963.

Yan, R. and Gao, R. (2009). Base wavelet selection for bearing vibration signal analysis. *International Journal of Wavelets, Multi-resolution, and Information Processing*, 7(4):411–426.

Zhang, Q. and Benveniste, A. (1992). Wavelet networks. *IEEE Transactions on Neural Networks*, 3(6):889–898.