Reducing image search time by improved BOVW with wavelet decomposition

Mohammed El Amin Kourtiche[1], Mohammed Beladgham[2], Abdelmalik, Taleb-Ahmed[3]
[1]Department of Computer Science, Laboratory of TIT, Tahri Mohammed University, Bechar, Algeria
[2]Department of Electrical Engineering, Laboratory of TIT, Tahri Mohammed University, Bechar, Algeria
[3]Laboratory of IEMN DOAE. UMR CNRS 852, University of Valenciennes, Valenciennes, France

Article Info

Article history:

Received Jun 15, 2020
Revised Aug 22, 2020
Accepted Sep 7, 2020

Keywords:

Bag of visual words
Image retrieval
Indexing
Wavelet decomposition

ABSTRACT

In the last decade, the bag of visual words (BOVW) has been used widely in image classification, image retrieval and has significantly improved the performance of CBIR system. In this paper we propose a new method to enhance BOVW using features obtained from wavelet decomposition in order to reduce computational costs in vocabulary construction and training time. We apply several level of wavelet decompositions and evaluate their impact on accuracy of the BOVW. We apply our method on MURA-v1.1 dataset and the experiments results confirm the performance of our approach.

Corresponding Author:

Mohammed El Amin Kourtiche
Department of Computer Science
Tahri Mohammed University
Independence road B.P 417, Bechar 08000, Algeria
Email: kourtiche@gmail.com

1. INTRODUCTION

Due to technological development and digital equipment, the image has become omnipresent in our daily lives, and as a result a hudge mass of images are stored every year, it is a very difficult task to retrieve similar images from such a huge database. The first studies focused on manual annotation which takes a lot of time and requires a large amount of manual processing in addition, it still subjective and may not correctly describe the content of the image so the Content Based Image Retrieval systems (CBIR) have appeared to fill the gaps in image searching area. CBIR systems are based on two phases the first is extraction of image features and indexing, the second phase is searching. The indexing techniques are employed to organize and effectively represent the contents of a database. As the images are interpreted as index vectors, then the search is done in the index database. The search process consists of searching for similar or near vectors by measuring the distance between the two vectors (usually the Euclidean distance) and using algorithms such as k-Nearest Neighbor. Among the approaches to search similar images or to classify an image to their own category we can cite the Bag of visual words (BOVW). BOVW has received much attention in the past decade and has been revisited and discussed in several publications. This paper presents a new approach to improve the performance of BOVW, for this way we will use the features generated by wavelet decomposition and use them in BOVW approach, which give us outperforms results in term of response time, storage capacity. The organization of this paper is as follows: Section 2 gives a brief overview of related work then Section 3 describes the BOVW method, Section 4 analyses wavelet decomposition process, our proposed approach is outlined in the Section 5, experimentation and results discussion in Section 6, and our conclusion is drawn in the final section.

2. RELATED WORK

Various systems have been introduced for CBIR systems using the extraction of low level features of image, such as color, shape and texture; recent techniques are based on keypoints extracted from regions-of-interest of the images which contain rich local information about the image[1], this keypoints can be detected using different detectors [2] and represented by several descriptors [3]. For indexing data and minimize the path to retrieve an image many tree structures have been introduced such as R-tree, R*-tree, and SR-tree, unfortunately the speed and accuracy of these algorithms degrade in high dimensional space, known as the curse of dimensionality, so several solutions have recently been proposed to solve this problem such as: Reduction of the size such as PCA (principal component analysis), Parallel architecture, Hashing method such as LSH (Locality Sensitive Hashing), Bag Of Visual Words (BOVW) which we describe it in detail in the next section.

Bag Of Visual Words has been revisited and discussed in several publications, in [1] the authors studied various representation choices such as vocabulary size, weighting, word selection and their impact to classification performance. In [4] The experiments confirm that the performance of a BOVW system can be greatly enhanced by taking the descriptors spatial distribution into account using the partitioning of the images with geometric tiling masks. In [5] exploit fuzzy clustering for codebook generation in combination with soft assignments and compared it with the traditional codebook approach using hard assignment. In [6] work, proposed an approach to incorporate spatial information in the BOVW on explicit global relationships among the spatial positions of visual words, inspired by the work of the [7] who proposed spatial pyramid match (SPM). In [8] proposed a new approach to revisited SDLC (Sorted Dominant Local Color), that divides the images into blocks, and generates a textual signature for each block and using weighting scheme based on the frequency of the visual words in the collection. To enhance SDLC another new S-BOVW mapping function, called Sorted Dominant Local Color and Texture (SDLCT) proposed in [9], this technique consists of a combination of representations based on both color as well as texture information [10] propose to add semantic information using eigenvectors of the image patches in order to extend the BOVW representation for image retrieval [11] Proposed a method for learning weighting schemes based on Genetic Programming to boost the performance of classification models relying on the BOVW.

3. BAG OF VISUAL WORDS

Use bag of visual words (BOVW) in image retrieval tasks was first proposed by [12] taking inspiration from the approach of bags of words indexing and searching of textual information, the texts are represented by sets of words from a vocabulary built from the corpus, therefore every text is represented by a histogram. This representation has proved to be very effective in image classification, object recognition and object detection. Use BOVW model to find similar images, to retrieve or to assign them to their own category can be divided into two major parts. The first part is the image representation and the second part consist to train a kernel method. As illustrated in Figure 1 the representation part can be divided into three steps:

The first step is the extraction of features which is the result of the detection of local points of interest (key points) and then describe them. The keypoints can be detected by various detectors [2] and described by different descriptors [3] like SIFT, PCA-SIFT, SURF, BRISK. After extracting feature vectors from each image in the dataset, the next step is to construct the dictionary (vocabulary or codebook), which is achieved by clustering the feature vectors obtained from all images of dataset in step1, using clustering algorithm as k-means or its variant. Each cluster centers (i.e, centroids) are treated as a visual word of the dictionary and the vocabulary size is the number of clusters. The last step in this part is vector quantization, used to quantify and represent each image in the dataset by a histogram of length k which refer to the number of clusters generated from k-means (visual vocabulary), where each local descriptor of dimension d from an image is assigned to the closest centroid, and the *i-th* value in the histogram is the frequency of the *i-th* visual word in the image. The histograms work as an indexing vocabulary, this means that the BOVW reduces the size of the image descriptors and provides a compressed representation of each image in the dataset. The second part of BOVW consists to train a classifier from labeled images based on the representation obtained from the first part. The most popular classifier is Support Vector Machines (SVM). SVM is flexible where the learning kernel can be varied according to the type of data used [13, 14] such as: linear, quadratic, Radial Basis Function (RBF), $\chi 2$, and EMD.

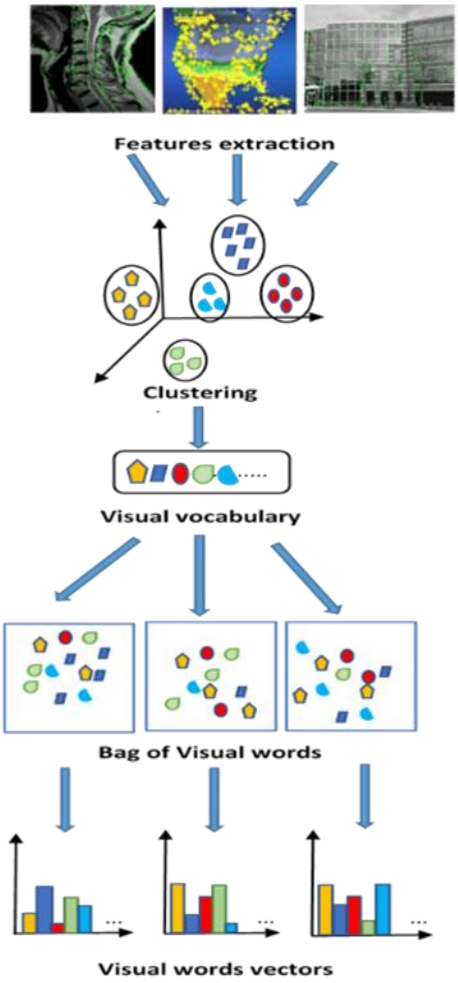


Figure 1. Image representation using bag of visual words

4. WAVELET AND GALL WAVELET

4.1. Wavelet decomposition

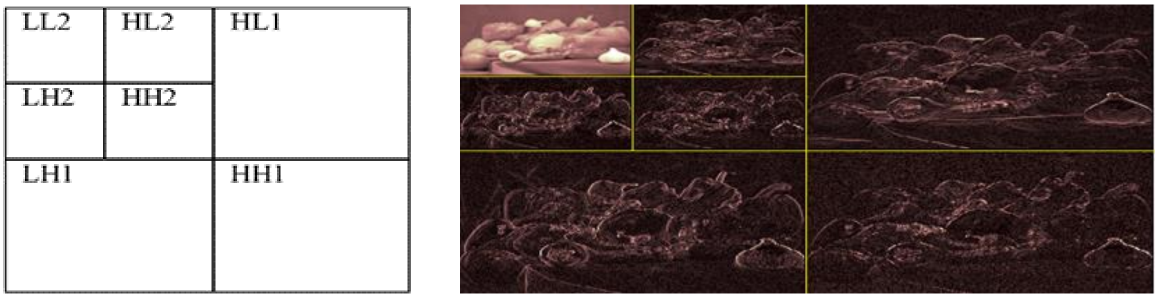
Wavelet decomposition has been widely used in image processing in various fields: biometric identification, compression, classification, image retrieval, image Watermarking [15-23] and has many advantages over Fourier transform. Wavelet transform is a well localized in both the time and frequency domain. Therefore, it may decompose a signal retaining the information of both domains. Wavelet transform decomposes a signal with a family of basis functions $\psi_{m,n}(x)$ obtained through translation and dilation [24] of a mother wavelet.

$$\psi_{m,n}(x) = 2^{-m/2}\psi(2^{-m}x - n)$$

where m and n are dilation and translation parameters.

The wavelet transformation on a 2D signal includes recursive filtering and sub-sampling [25]. The wavelet transforms can be computed by first performing 1D DWT (horizontally) on the rows. Then we will do the same 1D DWT on the columns (vertically) for both the low-pass and high-pass subband signals obtained from the horizontal analysis

At each level the signal is decomposed into four frequency sub-bands as shown in the Figure (a), LL called approximation, LH known as vertical details, HL called horizontal details, and HH known as diagonal details, where L denotes the low frequency and H denotes the high frequency. For the next decomposition level (i.e second level), we used this same process, however we take the approximation sub-image of the previous decomposition level (i.e first level) as shown in the image Figure (b).



(a) (b)

Figure 2. Wavelet decomposition of 2D-image signal (a). 2-level wavelet decomposition. (b). DWT of the peppers image over two levels

4.2. Biorthogonal wavelets

Biorthogonal wavelets 5/3 are part of the family of symmetric biorthogonal wavelets of Cohen-Daubechies-Feauveau (CDF). They are so called because the support width of their low-pass filters, detailed in Table 1, is p = 5 samples for analysis and p=3 for synthesis. In addition, they have $N = \widetilde{N} = 2$ zero moments. Due to their relative simplicity and the symmetry they offer, the 5/3 wavelets presented in Figure 3 are used enough in image coding. The wavelets of this family are also called Gall $(N, \widetilde{N})$, where N denotes the number of null moments of the analysis wavelet ψ and $\widetilde{N}$ its equivalent to synthesis. As for Daubechies wavelets, it is possible to show that Gall wavelets have minimal support for a given number of null moments $(N, \widetilde{N})$.

Table 1. Coefficients of symmetrical impulse responses of low-pass filters analysis $h_0[n]$ and synthesis $\tilde{h}_0[n]$ associated with Gall wavelets 5/3

n	$h_0[n]$	$\tilde{h}_0[n]$
0	1.06066017177982	0.70710678118655
1	0.35355339059327	0.35355339059327
2	-0.17677669529664	

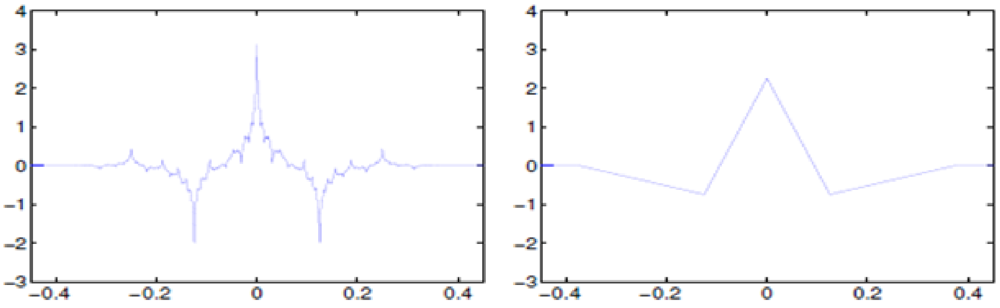


Figure 3. Gall wavelet 5/3 analysis ψ and its dual $\tilde{\psi}$.

5. OUR APPROACH

Our objective is to improve the response time of the BOVW by fast indexing the images contained in the dataset without having a significant decrease in accuracy when searching the convenient class for a query image, to this end our contribution consists to incorporate the Gall wavelet decomposition technique during the representation phase and evaluate their impact in accuracy and response time, Figure 2 illustrates our approach. Following is the detailed working procedure for the proposed approach:

a) The images of dataset are partitioned into two sets, one for training step and the other for testing. For each image in the training set we apply the wavelet decomposition. We take only the approximation part (sub-image) instead of taking the image in its totality.

b) Feature extraction of each sub-image is the next step, and is performed by the SURF method.

c) The construction of the vocabulary is obtained by clustering all vectors of previous step using KNN algorithm, the center of each cluster is called a visual word and the combination of visual words determines the dictionary.
d) Using the frequencies of each visual word in an image, we can represent every image by a histogram, that is used to train the classifier.
e) Training classifier using multiclass SVM: SVM is one of the most widely used classification models in the machine learning applications [26], it is commonly used in the classification of data especially in high dimensional feature spaces. In BOVW, the SVM classifier is trained using histograms of training images. SVM is used to establish an optimal hyperplane which separates the different classes of examples. SVM runs an algorithm that assists in finding the optimal hyperplane based on the training data. The optimal hyperplane is chosen such that the distance of the hyperplane from the nearest data point on either side is maximum. In our proposed method, the multiclass SVM classifier is used with linear kernel to classify the images dataset. Give the labeled training data to the SVM, it produces an optimal hyperplane, which then allows it to predict the class of a query image.

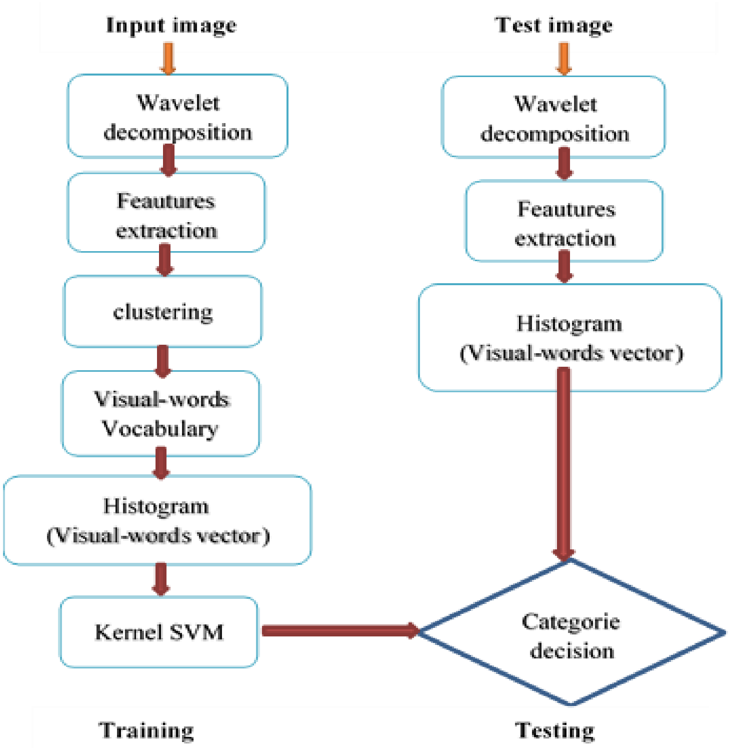


Figure 4. Bag of visual words pipeline for our approach

6. OUR EXPERIMENTS

The experiments were performed on medical Xray images MURA-v1.1, MURA is one of the largest public radiographic image datasets used in image retrieval tasks, classified by category according to the parts of the body examined (Finger, Forearm, Hand, Shoulder, Elbow, Humerus, Wrist) more details on this topic can be found in [27]. We aim to investigate the impact of using a wavelet decomposition in the Bag of Visual Words and evaluate different parameters: Vocabulary construction time, Training time, Storage and Accuracy. In our experiments we take three categories (Elbow, Hand, Shoulder) Figure 3 illustrates some samples.

The BOVW is flexible whereas different techniques can be used in each of its parts, so in our experiments, we chose SURF method to extract features and K-means for clustering them, we adopt SVM which is among the most widely used classifiers for BOVW. We fixed our size vocabulary on 500 visual words, for every experiment we use 30% images from each category to train the algorithm and 70% for testing phase. The experiments are tested on a system with Intel (R) Core (TM) i5- 7300U CPU 2.6 GHz and 8 GB RAM.

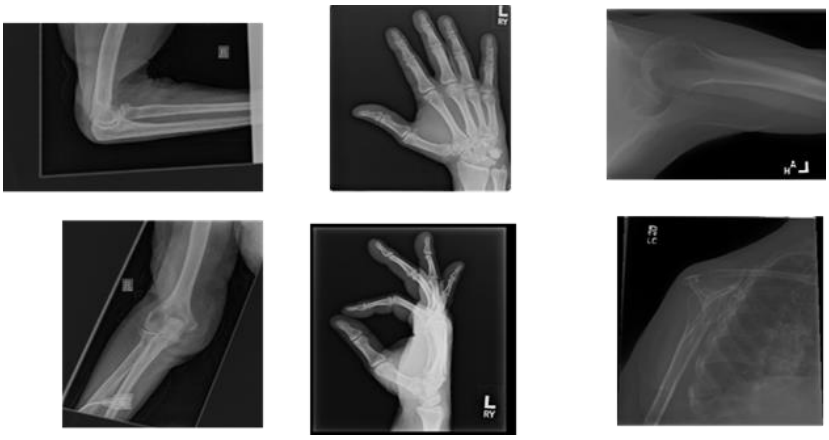


Figure 5. Sample images from MURA dataset

6.1. Discussion

The first set of analyses investigated between the original dataset and the decomposed dataset on level 1 is shown in Figure 4, we observe almost the same value of accuracy in the case of 3000, 4500, 6000 images. We note that in the traditional BOVW the machine cannot calculate the last two cases 7500 and 9000 images due to the exceeding of the capacity of the machine. In the representation phase we add the time of wavelet decomposition which represents 3% to 7% in this level (1) compared to the totality of time representation in level 1, Figure 5 shows a significant difference during the vocabulary construction phase. **Error! Reference source not found.**Figure 6 presents the training time, we note that the time in our approach in the case of 9000 is lower than the traditional BOVW in the case of 3000. From the histogram of Figure 7 we can see that the storage capacity of dataset of our approach represents 22% of the original dataset.

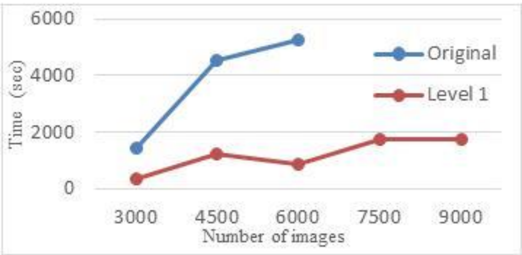


Figure 6. Accuracy

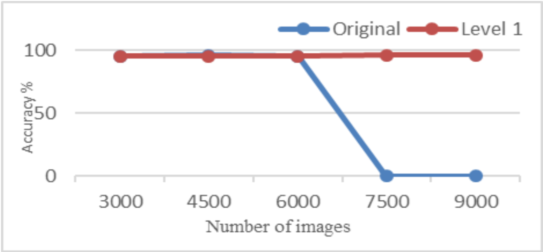


Figure 7. Vocabulary construction

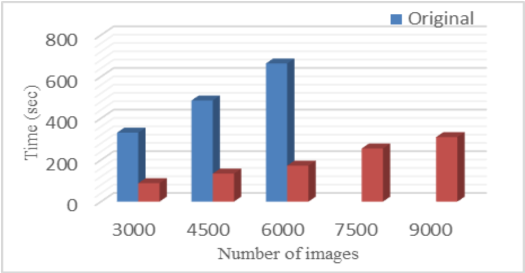


Figure 8. Kernel training

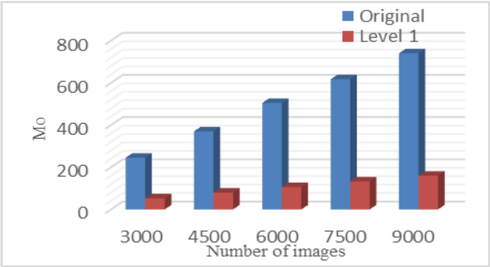


Figure 9. Storage

Our technique shows clearly has an advantage on vocabulary time construction and training time with minimum storage capacity compared to traditional one. In the following studies we want to find the suitable decomposition level which does not affect the accuracy rate with a reasonable response time, for this we vary in the decomposition level and the m, n size of dataset, the results are shown below. It is interesting to note that each time when we increase the level of wavelet decomposition, the response and training kernel times as well as the storage capacity are reduced. From the dictionary construction time as shown in the

Figure 10, we note that level 1 consumes more computation time whereas level 2 does not exceed 42% of the construction time of level 1 and is well clearly shown in all other cases. Figure shows that the accuracy at levels 4 and 5 is under 90%, and it is almost similar between levels 1 and 2 with a slight difference at level 3, so according to the graphs we conclude that the decomposition at level 2 allowed us to achieve almost the same accuracy as that original dataset with a considerable gain in time and storage medium. These tests revealed that wavelet decomposition in level 2 is the suitable solution for fast indexing and training BOVW. Kernel training and Storage as shown in Figure 11 and 12.

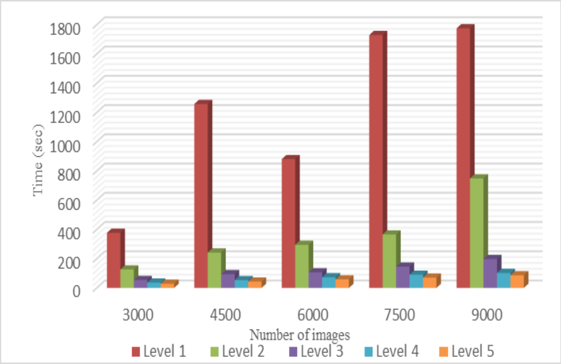


Figure 10. Vocabulary construction

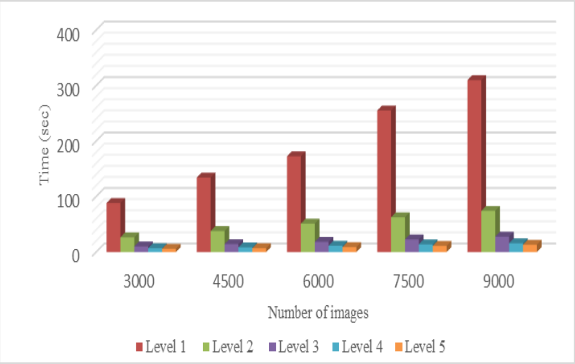


Figure 11. Kernel training

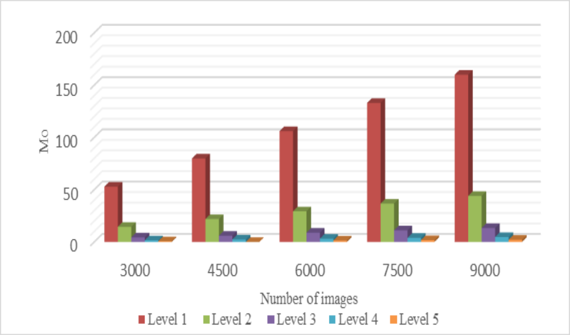


Figure 12. Storage

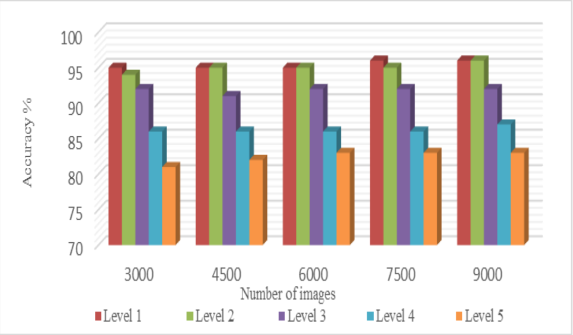


Figure 13. Accuracy

7. CONCLUSION

In this work, we propose a new method using wavelet decomposition in BOVW. As stated in the introduction our main aim is to reduce the time of indexing to retrieve and search relevant images. We studied also the impact of the wavelet decomposition level, our method allowed BOVW to reduce the number of the features vectors which impacts directly on the computational cost in BOVW and as expected our experiments prove that use wavelet decomposition in BOVW pipeline have a significant benefit in indexing time, kernel training and storage capacity.

REFERENCES

[1] J. Yang, Y.-G. Jiang, A. G. Hauptmann *et al.*, "Evaluating bag-of-visual-words representations in scene classification," *in Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval*, Augsburg, Bavaria, Germany, pp. 197-206, 2007.

[2] K. Mikolajczyk, and C. Schmid, "Scale & Affine Invariant Interest Point Detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63-86, 2004.

[3] K. Mikolajczyk, and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans Pattern Anal Mach Intell*, vol. 27, no. 10, pp. 1615-1630, 2005.

[4] V. Viitaniemi, and J. Laaksonen, "Spatial extensions to bag of visual words," *Proceedings of the ACM International Conference on Image and Video Retrieval*, pp. 1-8, 2009.

[5] M. Kogler, and M. Lux, "Bag of visual words revisited: an exploratory study on robust image retrieval exploiting fuzzy codebooks,"*Proceedings of the Tenth International Workshop on Multimedia Data Mining*, pp. 1-6, 2010.

[6] R. Khan, C. Barat, D. Muselet *et al.*, "Spatial orientations of visual word pairs to improve bag-of-visual-words model," *in: British Machine Vision Conference, BMVA*, 2012.

[7] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 2169-2178, 2006.

[8] J. M. dos Santos, E. S. de Moura, A. S. da Silva *et al.*, “A signature-based bag of visual words method for image indexing and search,” *Pattern Recognit Lett,* vol. 65, pp. 1-7, 2015.

[9] J. M. Dos Santos, E. S. De Moura, A. S. Da Silva *et al.*, “Color and texture applied to a signature-based bag of visual words method for image retrieval,” *Multimedia Tools and Applications,* vol. 76, no. 15, pp. 16855-16872, 2017.

[10] N. Bhattacharya, and J. Sil, "Image retrieval using extended bag-of-visual-words," I*nternational Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1969-1975, 2016.

[11] R. T. Ionescu, and M. Popescu, "Object recognition with the bag of visual words model," *Knowledge Transfer between Computer Vision and Text Mining: Similarity-based Learning Approaches*, pp. 99-132, Cham: Springer International Publishing, pp. 99-132, 2016.

[12] 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*, pp. 1470-1478, 2003.

[13] J. Zhang, M. Marszałek, S. Lazebnik *et al.*, “Local features and kernels for classification of texture and object categories: A comprehensive study,” *International Journal of Computer Vision,* vol. 73, no. 2, pp. 213-238, 2007.

[14] Y.-G. Jiang, C.-W. Ngo, and J. Yang, "Towards optimal bag-of-features for object categorization and semantic video retrieval," *Proceedings of the 6th ACM international conference on Image and video retrieval,* pp. 494-501, 2007.

[15] Y. Wang, X. Ding, R. Wang *et al.*, "Fusion-based underwater image enhancement by wavelet decomposition," *IEEE International Conference on Industrial Technology (ICIT)*, pp. 1013-1018, 2017.

[16] A. B. Said, I. Jemel, R. Ejbali *et al.*, "A hybrid approach for image classification based on sparse coding and wavelet decomposition," *IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA),* pp. 63-68, 2017.

[17] N. Qazi, and B. W. Wong, "Semantic based image retrieval through combined classifiers of deep neural network and wavelet decomposition of image signal," *Proceedings of the 9th EUROSIM Congress on Modelling and Simulation, EUROSIM,* no. 142, pp. 473-478, 2016.

[18] P. A. Hagargi, and D. Shubhangi, “Brain tumor MR image fusion using most dominant features extraction from wavelet and curvelet transforms,” *Brain,* vol. 5, no. 05, pp. 33-38, 2018.

[19] A. Setyono, and D. Setiadi, “Image Watermarking using Discrete Wavelet-Tchebichef Transform,” *Indones. J. Electr. Eng. Comput. Sci.,* vol. 16, no. 3, pp. 16-21, 2019.

[20] A. Vaish, S. Gautam, and M. Kumar, “A wavelet based approach for simultaneous compression and encryption of fused images,” *Journal of King Saud University-Computer and Information Sciences,* vol. 31, no. 2, pp. 208-217, 2019.

[21] C. Qin, Q. Zhou, F. Cao *et al.*, “Flexible lossy compression for selective encrypted image with image inpainting,” *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 29, no. 11, pp. 3341-3355, 2018.

[22] Z. Ye, H. Mohamadian, and Y. Ye, "Information measures for biometric identification via 2d discrete wavelet transform," *IEEE International Conference on Automation Science and Engineering*, pp. 835-840, 2007.

[23] P. Srivastava, and A. Khare, “Integration of wavelet transform, local binary patterns and moments for content-based image retrieval,” *Journal of Visual Communication and Image Representation,* vol. 42, pp. 78-103, 2017.

[24] M. Antonini, and M. Barlaud, “Image coding using wavelet transform," *IEEE Transaction on Image Processing,* vol. 1, no. 2, pp. 205-220, 1992.

[25] C.-W. Kok, and W.-S. Tam, *Digital Image Interpolation in Matlab*: John Wiley & Sons, 2019.

[26] O. Chapelle, P. Haffner, and V. N. Vapnik, “Support vector machines for histogram-based image classification,” *IEEE Trans Neural Netw,* vol. 10, no. 5, pp. 1055-1064, 1999.

[27] P. Rajpurkar, J. Irvin, A. Bagul *et al.*, “Mura: Large dataset for abnormality detection in musculoskeletal radiographs,” *arXiv preprint arXiv:1712.06957*, 2017.