Al-Nahrain Journal of Science

ANJS, Vol.24 (2), June, 2021, pp. 78-86

Modified Bag of Visual Words Model for Image Classification

Zainab N. Sultani* and Ban N. Dhannoon

Computer Science Department, College of Science, Al-Nahrain University, Baghdad, Iraq

Article's Information

Received:
28-05-2021
Accepted:
25-06-2021
Published:
27-06-2021

Keywords:
BoVW
Local features
SIFT
ORIB
Image classification

Abstract

Image classification is acknowledged as one of the most critical and challenging tasks in computer vision. The bag of visual words (BoVW) model has proven to be very efficient for image classification tasks since it can effectively represent distinctive image features in vector space. In this paper, BoVW using Scale-Invariant Feature Transform (SIFT) and Oriented Fast and Rotated BRIEF (ORB) descriptors are adapted for image classification. We propose a novel image classification system using image local feature information obtained from both SIFT and ORB local feature descriptors. As a result, the constructed SO-BoVW model presents highly discriminative features, enhancing the classification performance. Experiments on Caltech-101 and flowers dataset prove the effectiveness of the proposed method.

DOI: 10.22401/ANJS.24.2.11

*Corresponding author: zainab.namhabdula@nahrainuniv.edu.iq

1. Introduction

In the last decade, Bag of Visual Words (BoVW) features have been adopted in image classification or categorization tasks. This approach inspired by document representation techniques in text classification known as Bag of Words (BoW), where the text is represented as a vector of terms, with the same concept, the image is represented as a 1-Dimensional feature vector of an unordered bag of local feature descriptors, [1]. The principle of feature extraction is to obtain the most relevant features from the image data to obtain a sufficient and robust descriptor. Feature extraction is a crucial technique in the field of computer vision and image processing tasks like image classification. Feature extraction is a challenging subject as the features vary significantly due to several factors like noise, variations and scale, [2].

The reason for the success of using the BoVW model in image classification is the generic image keypoint description procedure by simply counting feature descriptors of an image. Keypoints are salient image patches that hold the image's local information. With supervised machine learning (ML) algorithms, such as k-Nearest Neighbour (kNN) [3], a category ML model is trained using a set of training images over the BoVW model representation. As various local image patches may represent parts of different objects represented in the same image. The BoVW model representation can describe a person, car, and even a landscape using a sufficient number of training images, [1].

Important features hold distinctive information and can differentiate one object from others. Local features represent the image patches, which are a small group of pixels. Low level descriptors that define an image with a feature vector using local level visual attributes such as scale-invariant feature transform (SIFT) [4], speeded-up robust feature (SURF) [5], local binary pattern (LBP), and Histogram of Oriented Gradient (HOG) [2] and Oriented Fast and Rotated BRIEF (ORB), [6].

Although various researches have lately been proposed to enhance the low-level feature descriptor, it remains an open study area. The extraction of the most valuable features is a vital step to improve the classification performance, [2].

In this paper, a new local feature descriptor (SO-BoVW) is represented by combining SIFT and ORB based on the BoVW model for image classification, where supervised ML k-Nearest Neighbour (kNN) is used to classify the images.

The rest of the paper is organized as follows: Section 2 provides a review of some related work. The proposed method is described in Section 3. Section 4 presents and discusses the experimental results. Finally, a conclusion is presented.

2. Related Work

Many techniques have been designed to extract image content characteristics from the visual data and use the extracted image content information for image recognition in response to the query image, [7].

In [8] the authors proposed an image classification model based on BoVW using four feature descriptors, SIFT, ORB, BRISK, and SURF. Caltech101 dataset was used, more specifically 261 images with three different labels; Airplanes, Helicopter and Motorbike. Their average accuracy with ORB, SIFT, BRISK, and SURF using kNN classifier was 57 %, 71 %, 62 %, and 77 %. Support Vector Machine (SVM) gave their best classification result using SURF descriptor, where the classification average accuracy was 85%.

Authors in [9] presented image feature extraction and classification using dangerous objects dataset. They proposed different models: BoVW using SURF as a feature descriptor and SVM classifier, HOG descriptor with SVM classifier, and CNN. Two thousand images are used where 1000 classified as images of knives while the other half without. BoVW presented very comparable results to CNN, where the accuracies were 84% and 87%, respectively.

Interferometric Synthetic Aperture Radar (InSAR) in [10] images were classified into eight classes using BoVW using two feature descriptors Gabor and Fractional Fourier Transform. kNN was used to classify the images using the first nearest neighbour and Euclidean distance. Four hundred images are used with only 4% for training the BoVW repeated 100 times. The overall accuracy is 81.79% which is obtained using Fractional Fourier Transform.

Global and local features are implemented in [11]. BoVW was built using LBP and ORB with global features to form a single vector applied on the Flowers dataset with different classifiers. Using the kNN classifier, their model has an accuracy of 19.4 %, while the Random Forest classifier was their best classification model; specifically, with BoVW, the accuracy was 64.13 %.

In [12], the authors conducted a BoVW model to classify images from the Caltech101 dataset. SIFT and kNN are used to build the model. Three classes are used (car_side, ship (ferry), and motorbike). Their model accuracy for the three classes was 90 %, 80 % and 80 %, respectively.

3. Proposed Methods

In this paper, image classification is conducted using a modified BoVW that combines two local feature descriptors SIFT and ORB (SO-BoVW), see Figure 1. First, the dataset is divided into training and testing sets, and then both SIFT and ORB are computed. For the training dataset, an unsupervised k-Means algorithm is utilized to create the set of the visual keywords feature vector for each descriptor. The size of the visual keywords is set to 100, 150, 200, and 250.

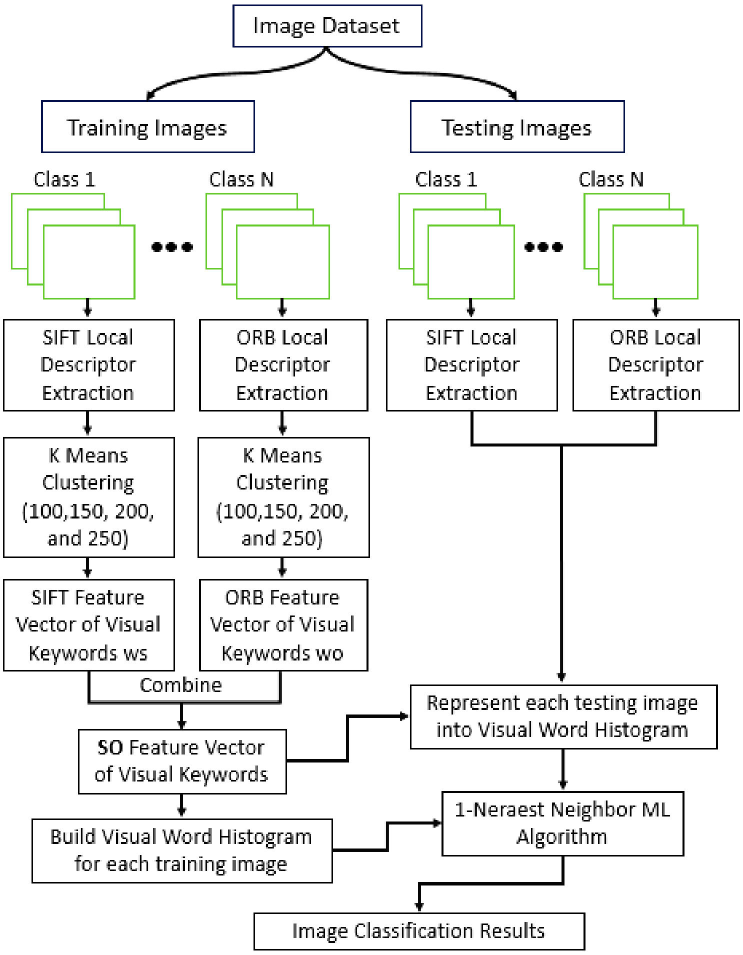


Image Dataset
Training Images
Testing Images
Class 1
Class N
Class 1
Class N
SIFT Local Descriptor Extraction
ORB Local Descriptor Extraction
SIFT Local Descriptor Extraction
ORB Local Descriptor Extraction
K Means Clustering (100,150, 200, and 250)
K Means Clustering (100,150, 200, and 250)
SIFT Feature Vector of Visual Keywords ws
ORB Feature Vector of Visual Keywords wo
Combine
SO Feature Vector of Visual Keywords
Represent each testing image into Visual Word Histogram
Build Visual Word Histogram for each training image
1-Neraest Neighbor ML Algorithm
Image Classification Results

Figure 1. Block diagram of the proposed SO-BoVW for image classification.

Then the new descriptor will be formed by combining (concatenating) both vectors. Each training image is represented as a frequency/histogram of all the visual words (SO). Then a data matrix of all training images is generated, see Figure 2. Then for each testing image, a visual keyword histogram vector is generated based on the SO feature vector. Finally, a 1NN is used to classify each testing image by computing the distance with all the training images and assign a label to the nearest distance. BoVW, SIFT, ORB, and kNN are explained in the following subsections.

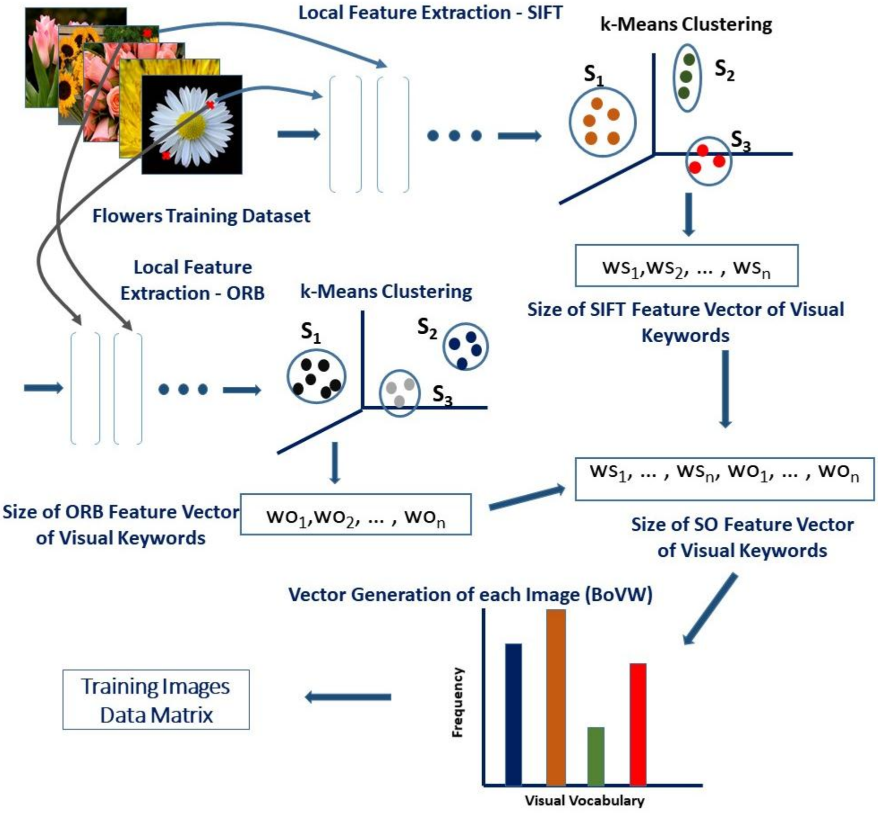

Local Feature Extraction - SIFT
k-Means Clustering
S_1
S_2
S_3
Flowers Training Dataset
ws_1,ws_2, ... , ws_n
Size of SIFT Feature Vector of Visual Keywords
Local Feature Extraction - ORB
k-Means Clustering
S_1
S_2
S_3
Size of ORB Feature Vector of Visual Keywords
wo_1,wo_2, ... , wo_n
ws_1, ... , ws_n, wo_1, ... , wo_n
Size of SO Feature Vector of Visual Keywords
Vector Generation of each Image (BoVW)
Frequency
Visual Vocabulary
Training Images Data Matrix

Figure 2. SO-BoVW Model framework.

3.1 Bag of visual words model

The BoW method was adopted for text retrieval/classification; then, it's expanded to computer vision. In text classification models, each text/document is expressed by term frequency. Usually, this includes all keywords (terms) from all documents by forming a vocabulary. The vocabulary may dismiss specific noninformative terms such as stop words, and it converts the terms to their base form. A text document is then represented by a sparse vector where each dimension represents a term in the vocabulary. The feature's value is the number of times the term appears in the document. The BoW representation is an order less collection of vocabulary terms. An image can be represented as a histogram (frequency) distribution of visual words, regardless of their spatial position in the image. Words can be easily extracted from a text document; however, visual words are more challenging to describe. Typically, local image features extracted at specific regions (patches) of interest (distinctive) are utilized to describe visual words. Local features from images are assigned to the closest word in the vocabulary. Then BoVW vector is built by counting the frequency of local features (from vocabulary) presented in the image [1]. The following steps summarize BoVW image representation steps [1,10]:

Step 1. Local feature detection and extraction: The images are divided into patches to detect the local features or keypoints. These detected feature regions are local patches in the image. Then feature descriptors are utilized to represent the keypoint by using the local neighbourhood. The most known descriptor is the SIFT, where eight gradients orientation with windows size (patch) 4×4 is used to form a feature vector of 128 dimensions.

Step 2. Dictionary (Vocabulary) Generation: Extraction of Local features over a large set of training images can sometimes result in many features with insignificant variations. The number of feature descriptors is reduced by vector quantization (VQ) approaches to build a dictionary (vocabulary) that provides some invariance to minor changes between features and, at the same time, decreases computational complexity. Most BoVW models implement unsupervised ML k-means to cluster the descriptors of a training image dataset into k dictionary words.

Step 3. BoVW feature vector generation: When the clusters' center are generated and the visual word dictionary is learned, they represent each image in the dataset by specifying all features descriptors of each image to the most similar dictionary feature vector. For each image, a frequency histogram of the visual word vector is generated, usually by applying the nearest neighbour search using distance measurement. The achieved frequency distribution is referred to as BoVW.

3.2 Scale Invariant Feature Transform (SIFT)

The SIFT is a local feature detection and description algorithm used to detect distinctive points invariant to image rotation and scaling. SIFT algorithm can be summarized in 4 steps as shown in Figure 3 [13] and explained as following [12,14,15]:

Step1. Determine approximate keypoints location and scale: in this step, the interest points' (keypoints) location and scale are defined using Difference of Gaussian (DoG) with different scale values.

Step2. Keypoint localization: after DoG, each pixel is compared with 26 pixels, eight neighbourhood pixels, in addition to 9 pixels in the upper and lower scales. Low contrast and edge points are eliminated using Hessian matrix.

Step3. Orientation assignment: to specify the dominant orientation histograms of gradient orientations are utilized. The 360 degrees of orientation are divided into 36 bins. For example, 20.5 degrees is assigned to 20-29 bin. This procedure is applied to all the pixels which are keypoint description neighbourhood; then the histogram will contain a peak which will be the assigned orientation, for example, if 20-29 bin is the highest point, then the keypoint is assigned the third bin (orientation 3).

Step4. Keypoint description: To generate a unique description (feature vector), a window of 16x16 around each keypoint is created then each cell is divided into window of 4×4 where it will have eight directions. A total of 128 (4×4×8) dimension vector is generated for each keypoint.

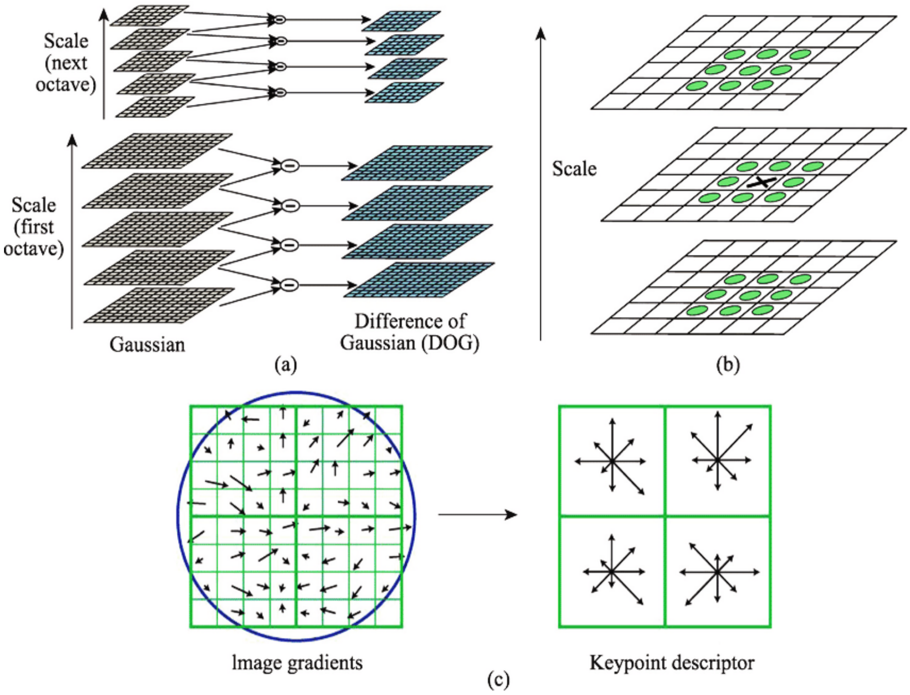


Figure 3. SIFT descriptor framework. (a) DoG, (b) Keypoint localization, (c) Orientation assignment and keypoint description.

3.3 Oriented Fast and Rotated BRIEF (ORB)

The Oriented FAST and Rotated BRIEF (ORB) algorithm was designed by Rublee et al in 2011. ORB is a blending of Features from Accelerated Segment Test (FAST) key point extractor and Binary Robust Independent Elementary Features (BRIEF) descriptor. ORB local feature descriptor algorithm has two steps: feature point extraction and feature point descriptors generation.

The ORB algorithm uses the improved (FAST) algorithm for feature point extraction. The main concept is to detect a corner point by exploring the difference between a pixel and its neighbourhood; if the difference is high, it's probably a corner point. The FAST corner points have both scale and rotation invariance. To enhance the rotation, invariance moments are computed. The ORB algorithm uses the improved (BRIEF) algorithm after the oriented FAST feature points are extracted to calculate the point descriptor. BRIEF is a binary vector descriptor (0,1). An ORB uses the Steer BRIEF since BRIEF doesn't have rotational invariance [16,17].

3.4 k-Nearest Neighbour Classifier (kNN)

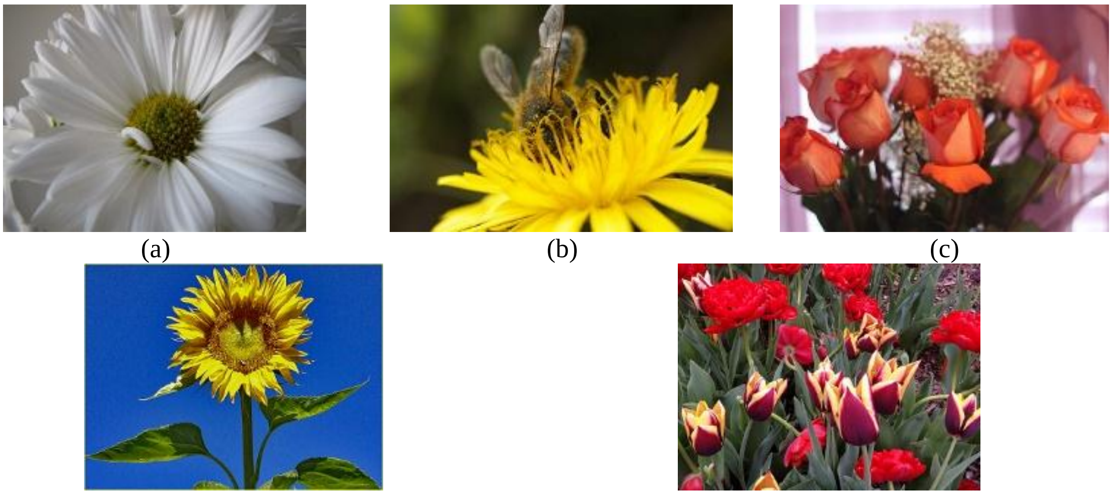
After the data matrix of the training images is generated using both SIFT and ORB descriptors, the classification phase can be conducted. First, the kNN model doesn't require a training phase since kNN is a lazy classifier, therefore for each testing image, the SIFT and ORB descriptors are extracted, and then the feature vector is generated using SO visual keywords. In order to classify the testing image, Euclidean distance is used to calculate the dissimilarity matrix with all training images in the data matrix. After sorting the distances, a training image with minimum distance is selected (k = 1; first nearest neighbour), and its label is assigned to the testing image.

4. Experimental Results and Discussion

In our experiments, the SO-BoVW model are designed for two datasets, Flowers [18] and Caltech101 [19] datasets. Hold-out Cross-Validation is used to divide the datasets into 70% training and 30% testing. Flowers dataset contains 3670 images in total, divided into 5 classes; see Figure 4. The training set contains 2567 images while 1103 testing images, taking into account the distribution of images among the 5 classes, see Table 1.

Table 1. Flowers dataset images per classes (training and testing).

Class	#of training images	#of testing images
Daisy	443	190
Dandelion	628	270
Roses	448	193
Sunflowers	489	210
Tulips	559	240
Total Number	2567	1103



(a) (b) (c)

(d) (e)

Figure 4. Flowers dataset sample (a) Daisy, (b) Dandelion, (c) Roses, (d) Sunflowers, (e) Tulips.

Caltech101 contains 101 different scenes; however, in this paper, 6 classes were chosen (Airplanes, Car Side, Chair, Cup, Helicopter, and Motorbikes), see Figure. 5. The number of images in total is 1928; using hold out, the total number of images in training and testing is divided among 6 classes, as shown in Table 2.

Table 2. Caltech10 Dataset images per 6 classes (training and testing)

Class	#of training images	#of testing images
Airplanes	560	240
Car_side	86	37
Chair	43	19
Cup	39	18
Helicopter	61	27
Motorbikes	558	240
Total Number	1347	581

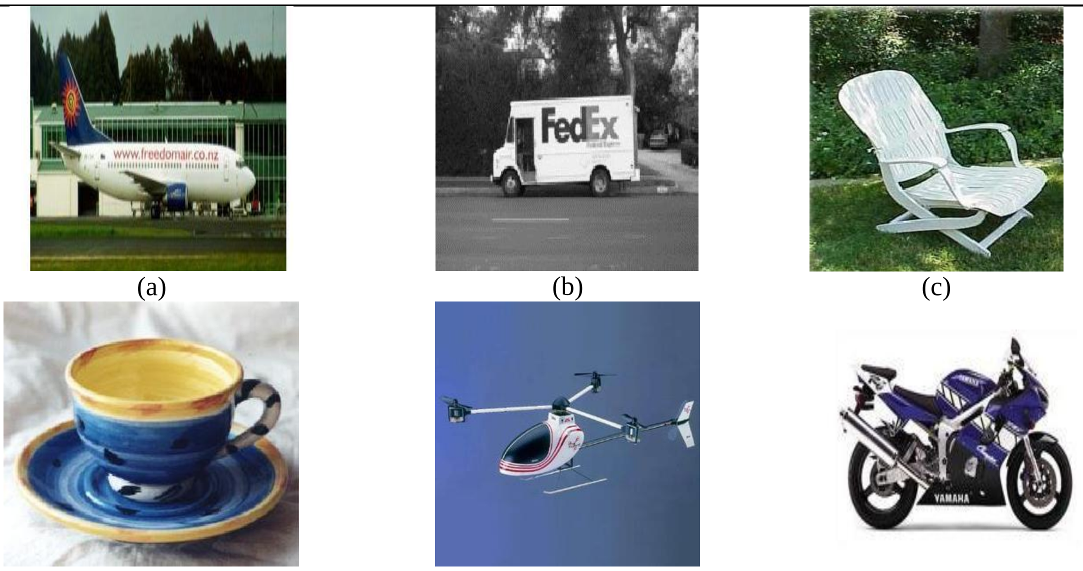

(a) (b) (c)
(d) (e) (f)

Figure 5. Caltech101 dataset sample, (a) Airplanes, (b) Car_side, (c) Chair, (d) Cup, (e) Helicopter (f) Motorbikes**.**

Three different BoVW models are presented: SIFT-BoVW, ORB-BoVW, and SO-BoVW. To demonstrate the effect of the feature vector size on the accuracy result, four values are implemented 100, 150, 200, and 250 for each model. First Flowers dataset with 5 classes accuracy results are recorded using ORB-BoVW and SIFT-BoVW models. Comparing the Average Accuracy of Tables 3 and 4 clearly shows that the SIFT features are better than ORB. But using the hybrid SO-BoVW is the best amongst the three models, see Figure. 6. In SO-BoVW, SIFT and ORB descriptor vectors are concatenated to form the final vector; therefore, in Tables 5 and 8, when the vector size is set to 100, it means 100 SIFT and 100 ORB. The flowers dataset is balanced where classes have an approximately similar number of images. In Table 3, the class-based accuracy is between 20%-50%. However, in Tables 4 and 5, the class-based accuracy increased to approximately 35%-60%. The best feature vector size is 200 for the three models. It can be said that the accuracy for the 5 classes is roughly close to each other.

Table 3. Flowers dataset class-based and average accuracy using ORB-BoVW.

	100	**150**	**200**	**250**
Daisy	%31.0526	%31.0526	%31.05263	%21.0526
Dandelion	%44.0740	%47.4074	%44.0740	%41.1111
Roses	%26.9430	%25.3886	%17.09844	%25.3886
Sunflowers	%35.2380	%39.5238	%30.9523	%33.8095
Tulips	%35.8333	%32.9166	%40.4166	%39.5833
Average Accuracy	%35.3581	%36.0834	%33.8168	%33.1822

Table 4. Flowers dataset class-based and average accuracy using SIFT-BoVW.

	100	**150**	**200**	**250**
Daisy	%47.8947	%40.5263	%44.2105	%45.2631
Dandelion	%54.4444	%58.1481	%62.9629	%55.9259
Roses	%35.2331	%36.2694	%39.3782	%35.7512
Sunflowers	%48.5714	%50.4761	%52.3809	%52.3809
Tulips	%39.5833	%37.9166	%41.6666	%45.0000
Average Accuracy	%45.6029	%45.4215	%48.9573	%47.5067

Table 5. Flowers dataset class-based and average accuracy using SO-BoVW.

	SIFT100&ORB100	**SIFT150&ORB150**	**SIFT200&ORB200**	**SIFT250&ORB250**
Daisy	%37.8947	%45.2631	%43.6842	%42.1052
Dandelion	%56.6666	%55.9259	%58.5185	%58.5185
Roses	%43.5233	%35.7512	%44.0414	%40.9326
Sunflowers	%53.3333	%52.3809	%50.9523	%55.2380
Tulips	%39.5833	%45.0000	%45.0	%40.0000
Average Accuracy	%46.78150	%47.5067	%49.0480	%47.9601

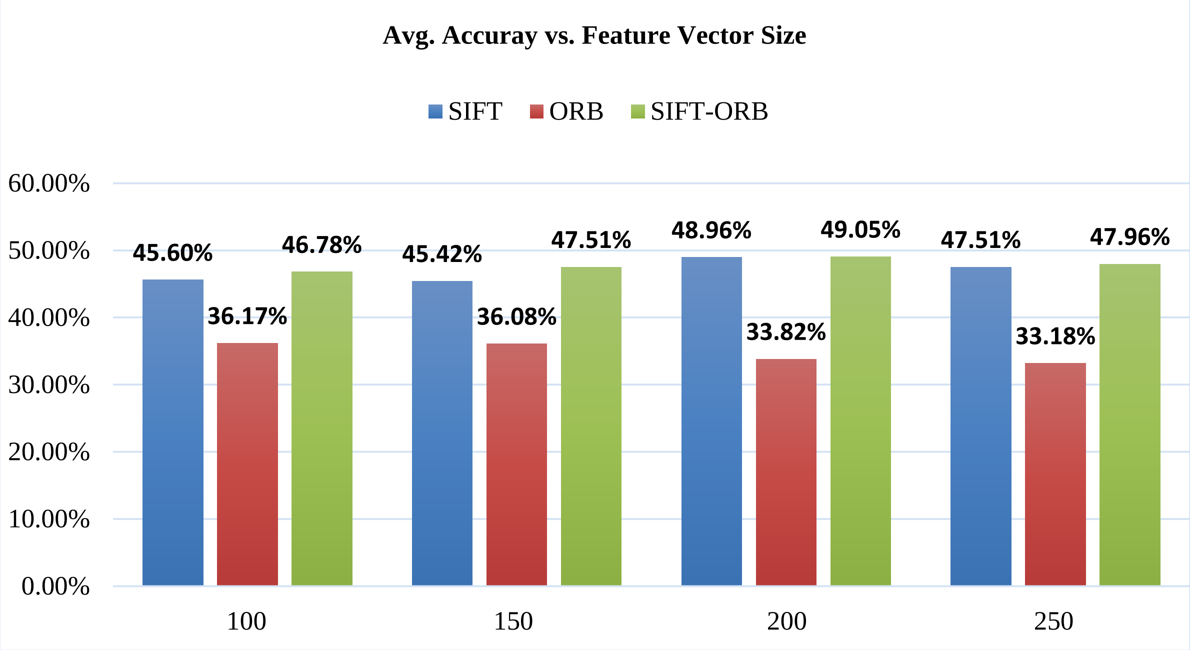


Figure 6. Accuracy results for SIFT, ORB and SO-BoVW using flowers dataset.

Second Caltech101 dataset with 6 classes' accuracy results are recorded using ORB-BoVW and SIFT-BoVW and SO-BoVW models, see Tables 6, 7, and 8. This dataset doesn't have an equal size of images among classes; therefore, it is an unbalanced dataset. The average accuracy using SIFT is better than ORB; however, the class-based accuracy is diverse; using SIFT provided better accuracy in "airplanes" and "car_side," but ORB gave better results in classifying "cup" and "motorbikes" classes. Also, as in the Flowers dataset, it is noticed that feature size of 200 was also the best among the four cases. The proposed SO-BoVW model presented the best results in terms of average accuracy, see Figure. 7, and for the majority of class-based accuracy.

Table 6. Caltech101 dataset class-based and average accuracy using ORB-BoVW.

	100	**150**	**200**	**250**
Airplanes	%77.5	%79.1666	%83.3333	%79.1666
Car_side	%45.9459	%43.2432	%21.6216	%29.7297
Chair	%5.2631	%5.2631	%0.0	%5.2631
Cup	%16.6666	%11.1111	%5.5555	%16.6666
Helicopter	%3.7037	%0.0	%0.0	%7.4074
Motorbikes	%92.0833	%92.5	%94.5833	%90.4166
Average Accuracy	%73.8382	%74.1824	%75.0430	%72.9776

Table 7. Caltech101 dataset class-based and average accuracy using SIFT-BoVW.

	100	150	200	250
Airplanes	%85.0	2	%88.33333	%80.8333
Car_side	%72.9729	%70.2702	%67.5675	%67.5675
Chair	%15.7894	%26.3157	%21.0526	%26.3157
Cup	%0	%0	%11.1111	%5.5555
Helicopter	%18.5185	%18.5185	%22.2222	%29.6296
Motorbikes	%90.0	%87.0833	%92.0833	%88.3333
Average Accuracy	%78.3132	%76.9363	%80.8950	%76.5920

Table 8. Caltech101 Dataset class-based and average accuracy using SO-BoVW.

	SIFT100&ORB100	SIFT150&ORB150	SIFT200&ORB200	SIFT250&ORB250
Airplanes	%85.0	%87.5	%87.9166	%86.6666
Car_side	%78.3783	%62.16216	%67.5675	%43.2432
Chair	%15.7894	%15.7894	%21.0526	%21.0526
Cup	%16.6666	%11.1111	%16.6666	%11.1111
Helicopter	%7.4074	%22.2222	%14.8148	%7.4074
Motorbikes	%99.1666	%97.0833	%99.1666	%97.9166
Average Accuracy	%82.4440	%82.0998	%83.4767	%80.3786

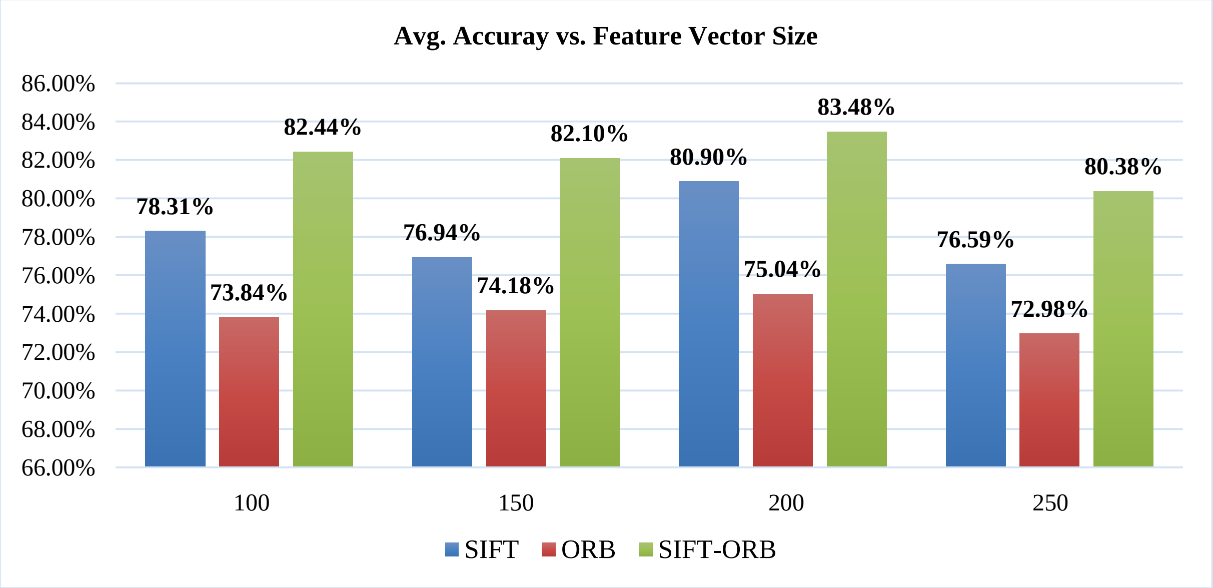


Figure 7. Accuracy results for SIFT, ORB and SO-BoVW using Caltech101 dataset.

Table 9 presents the accuracy results obtained by the proposed model descriptor SO-BoVW and other related works that used the Caltech101 dataset [8,12]. As shown in this table, the proposed model outperforms the model in [8], which uses either SIFT or ORB. But the best classification accuracy was reached by [12], where they have only used three classes, (Car_side, Ship, and Motorbikes) however in this paper, six classes were chosen (Airplanes, Car_side, Chair, Cup, Helicopter and Motorbikes) therefore different and versatile keypoints are generated.

Table 9. Comparison of average accuracy rates.

Methods	Dataset	Local Descriptors	Classifier	Accuracy
SO-BoVW	Caltech101	SIFT&ORB	kNN	83.48%
[8]	Caltech101	ORB SIFT	kNN	57% 71%
[12]	Caltech101	SIFT	kNN	86.6%

5. Conclusion

In this paper, a modified BoVW model is suggested. Local feature descriptors, SIFT and ORB are used as one feature vector to represent the images as a data matrix for classification. Two dataset are utilized to assess the proposed method. Four different feature vector sizes are used to select the best size. It was concluded that 200 was the best. However, more feature vector size can be used to study the effect more precisely. A small number of images per class has an impact on their class-based classification accuracy in all models. Since the flowers dataset is balanced, the class-based accuracy was close to each other, but in the Caltech101 dataset, there is a considerable gap between the number of images in classes, so the best accuracy was for the class with the highest number of images. It can be concluded that SIFT performs better than ORB in most scenarios, and the proposed SO-BoVW is the best in terms of average accuracy and for the majority of the class-based.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Hentschel C. and Sack H.; “Does one size really fit all? Evaluating classifiers in bag-of-visual-words classification”, ACM Int. Conf. Proceeding Ser., 16-19-Sept, 2014. doi: 10.1145/2637748.2638424.

[2] Kabbai L.; Abdellaoui M. and Douik A.; “Image classification by combining local and global features”, Vis. Comput., 35(5), 679-693, 2019. doi: 10.1007/s00371-018-1503-0.

[3] Altman N. S.; “An introduction to kernel and nearest-neighbor nonparametric regression”, Am. Stat., 46(3), 175-185, 1992. doi: 10.1080/00031305.1992.10475879.

[4] Lowe D. G.; “Distinctive image features from scale-invariant keypoints”, Int. J. Comput. Vis., 60(2), 91-110, 2004. doi: 10.1023/B:VISI.0000029664.99615.94.

[5] Bay H.; Tuytelaars T. and Gool L. V.; “LNCS 3951 - SURF: Speeded up robust features”, Comput. Vision-ECCV 2006, 404-417, 2006, [Online]. Available: http://link.springer.com/chapter/10.1007/11744023_32.

[6] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” Proc. IEEE Int. Conf. Comput. Vis., 2564-2571, 2011, doi: 10.1109/ICCV.2011.6126544.

[7] Verma V.; “Image Retrieval and classification using local feature vectors”, June, 2014. [Online]. Available: http://arxiv.org/abs/1409.0749.

[8] Dave M. and Israni D.; “Evaluating classifiers and feature detectors for image”, November, 2019.

[9] Kibria S. B. and Hasan M. S.; “An analysis of Feature extraction and classification algorithms for dangerous object detection”, 2nd Int. Conf. Electr. Electron. Eng. ICEEE 2017, 1-4, 2018. doi: 10.1109/CEEE.2017.8412846.

[10] Cagatay N. D. and Datcu M.; “Bag-of-visual-words model for classification of interferometric SAR images”, Proc. Eur. Conf. Synth. Aperture Radar, EUSAR, 243-246, 2016.

[11] Raikar P. and Joshi S. M.; “Efficiency comparison of supervised and unsupervised classifier on content based classification using shape, color, texture”, 2020 Int. Conf. Emerg. Technol. INCET 2020, 1-7, 2020. doi: 10.1109/INCET49848.2020.9154016.

[12] Karim A. A. A. and Sameer R. A.; “Image Classification using bag of visual words (BoVW)”, Al-Nahrain J. Sci., 21(4), 76-82, 2018. doi: 10.22401/anjs.21.4.11.

[13] Lyu W.; Zhou Z.; Chen L. and Zhou Y.; “A survey on image and video stitching”, Virtual Real. Intell. Hardw., 1(1), 55-83, 2019. doi: 10.3724/sp.j.2096-5796.2018.0008.

[14] Yi K. M.; Verdie Y.; Fua P. and Lepetit V.; “Learning to assign orientations to feature points”, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2016-Decem, 107-116, 2016. doi: 10.1109/CVPR.2016.19.

[15] Gv2, “Histograms of Orientation”, Statistics (Ber), 28, 2009.

[16] Luo C.; Yang W.; Huang P. and Zhou J.; “Overview of Image matching based on ORB algorithm”, J. Phys. Conf. Ser., 1237(3), 2019. doi: 10.1088/1742-6596/1237/3/032020.

[17] Karami E.; Prasad S. and Shehata M.; “Image matching using SIFT, SURF, BRIEF and ORB: Performance comparison for distorted images”, arXiv, 2017.

[18] Tung, K, "Flowers Dataset", https://doi.org/10.7910/DVN/1ECTVN. Harvard Dataverse, 8, 2020.

[19] Fei-Fei L.; Fergus R. and Perona P.; "Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories", IEEE. CVPR 2004, Workshop on Generative-Model Based Vision, 2004.