Computer Science and Information Systems 15(3):615–633 https://doi.org/10.2298/CSIS180105025Z

Intelligent Image Classification-Based on Spatial Weighted Histograms of Concentric Circles

Bushra Zafar[1], Rehan Ashraf[1], Nouman Ali[2], Mudassar Ahmed[1], Sohail Jabbar[1], Kashif Naseer[3], Awais Ahmad[4], and Gwanggil Jeon[5,6]

[1] Department of Computer Science
National Textile University, Faisalabad, Pakistan
bkgcuf@gmail.com,{rehan,mudassar}@ntu.edu.pk, sjabbar.research@gmail.com
[2] Department of Software Engineering
Mirpur University of Science and Technology, Mirpur, Azad-Kashmir, Pakistan
nouman.se@must.edu.pk
[3] Department of Computer Engineering
Bahria University, Islamabad
kashifnaseer85@yahoo.com
[4] Department of Computer Science
Bahria University, Islamabad
aahmad.marwat@gmail.com
[5] School of Electronic Engineering
Xidian University, China
[6] Department of Embedded Systems Engineering
Incheon National University, Korea
ggjeon@gmail.com

Abstract. As digital images play a vital role in multimedia content, the automatic classification of images is an open research problem. The Bag of Visual Words (BoVW) model is used for image classification, retrieval and object recognition problems. In the BoVW model, a histogram of visual words is computed without considering the spatial layout of the 2-D image space. The performance of BoVW suffers due to a lack of information about spatial details of an image. Spatial Pyramid Matching (SPM) is a popular technique that computes the spatial layout of the 2-D image space. However, SPM is not rotation-invariant and does not allow a change in pose and view point, and it represents the image in a very high dimensional space. In this paper, the spatial contents of an image are added and the rotations are dealt with efficiently, as compared to approaches that incorporate spatial contents. The spatial information is added by constructing the histogram of circles, while rotations are dealt with by using concentric circles. A weighed scheme is applied to represent the image in the form of a histogram of visual words. Extensive evaluation of benchmark datasets and the comparison with recent classification models demonstrate the effectiveness of the proposed approach. The proposed representation outperforms the state-of-the-art methods in terms of classification accuracy.

Keywords: Image Classification, Bag of Visual Words, Support Vector Machine, Weighted Histograms of Concentric Circles.

1. Introduction

Due to an increase in the volume of image contents in recent years, image classification has gained considerable attention in industry and academia [20]. The main objective in any classification-based approach is to classify the image according to the available category [39]. Variations in scale, spatial layout and illumination make image classification a challenging task [1]. In the past, the global color, texture and shape features have been used for image classification as global features are considered simple to implement; however their performance is lower.

The BoVW model is used for image classification, object recognition and image retrieval problems [24]. The visual information is extracted from the image by using local features and an order-less histogram of visual words is used for image representation [29]. In the approach, an image is treated like a document that consists of words. A group of training images is used to represent the images patches, which are known as a codebook or visual vocabulary . The steps used in the BoVW model are (i) feature extraction using a key-point detector (ii) codebook construction (iii) representation of the image as an order-less histogram of visual words [29, 23]. The BoVW ignores the spatial layout among the local features. yet according to research, this spatial information is beneficial for image classification-based problems [1, 18].

Two approaches are proposed to address this problem: (i) the geometric relationship among visual word co-occurrence (also known as visual word co-occurrence), and (ii) the division of an image into different regions for the construction of histogram of visual words [1, 4, 16]. Due to the complex relationship among visual words , the approaches based on visual word co-occurrences are reported to be computationally expensive [1, 4]. Among the techniques which divide the image into different regions, Spatial Pyramid Matching (SPM) is considered the most popular; it divides the images into rectangular regions and computes a histogram of visual words from each of the sub-regions.

SPM is considered an extension of the order-less BoVW and it addresses the shortcoming of a lack of spatial information [5, 37, 6]. However, if the objects in the images are flipped or rotated, the discrimination power of SPM decreases [37, 15]. Zhao *et al.* [37] propose a Concentric Circle-structured Multiscale BoVW model (CCM-BOVW) to incorporate rotation-invariant spatial layout information of images into the BoVW model for land-use scene classification. As a first step, CCM-BOVW constructs multiple resolution images and extracts multiple local features from all of the resolution images. This is followed by the visual vocabulary construction step, after which each image is partitioned into subregions by a set of concentric circles. Each subregion is represented separately by a histogram of visual words, and then the histograms from each subregion of an image are concatenated. To create the final histogram representation for an image, the histograms of different resolution images are combined.

According to Karmakar *et al.* [15], the construction of histograms over the concentric rectangles makes it possible to deal with image rotations and changes in view point. In this paper, we aim to address this problem by proposing a novel images representation that is based on histograms of concentric weighted circles. Consider the example of two cases that are shown in Fig. 1.

The upper region of Fig. 1 represents the idea of histograms constructed over the concentric rectangular regions [15], while the lower region represents the histograms con-

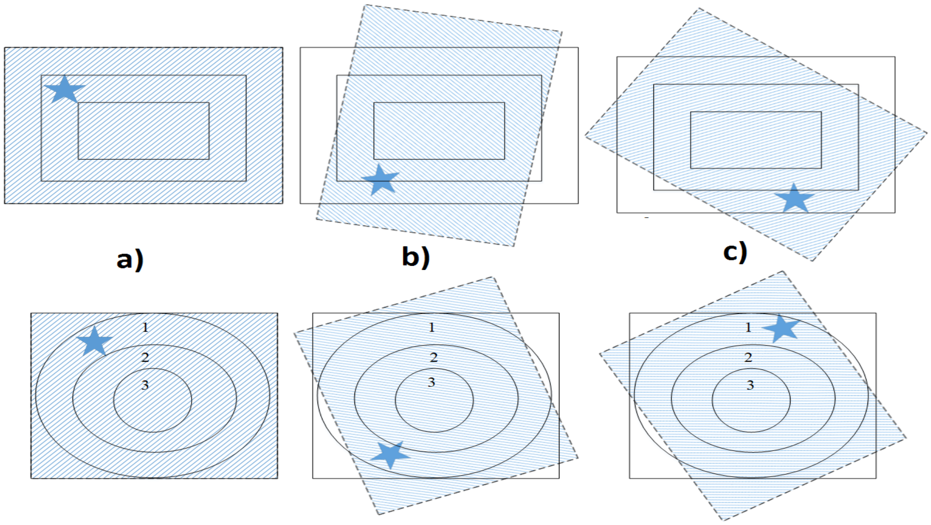


Fig. 1. A representation of how the proposed approach based on concentric circles is more discriminating than the concentric rectangles

structed by using the proposed approach based on weighted concentric circles. In the figures, we can see that

(a) the image is not rotated.
(b) the image is rotated by 110 degrees and
(c) the image is rotated by 300 degrees.

The visual words using rectangle coordinates are not equidistant from the center, so rotation invariance is not achieved, while in the circle, all points are equidistant from the center so it achieves rotation invariance (as compared to [15]).

The proposed method exploits the concentric circle strategy to improve the BoVW model. The proposed approach is different from the previously published work [37] as a weighted circle scheme is applied to represent the image in the form of a histogram of visual words. The higher level segments incorporate more spatial information compared to lower level segments and hence are assigned more weights. In addition to this, we extract Scale Invariant Feature Transform (SIFT) features on multiple scales using a dense grid [23] and apply a different kernel technique for image classification [32]. The main contributions of this research are

(i) the addition of spatial information to BoVW histogram representation.
(ii) a rotation-invariant approach to image representation.

The rest of the paper is organized as follows: the following section is a literature review. Section 3 provides an overview of the BoVW model and presents our proposed approach. Section 4 provides a discussion of the results obtained using the method on two benchmark datasets, as well as a comparison with the other state-of-the-art approaches. The last section concludes the paper and points towards future directions of research.

2. Related Work

A major limitation of the BoVW model is the fact that it ignores spatial information [30, 16]. Despite this fact, BoVW exhibits a high discriminative power and has shown excellent results in image classification [13, 12, 36]. However, numerous research studies demonstrated that the performance can be improved by incorporating the missing spatial information. Broadly, the related work can be classified into two groups [3, 17]. The first group comprises methods that encode relationships or co-occurrence of visual words [16, 26]. Khan *et al.* [16] made a notable contribution to this domain and incorporated global spatial information in the BoVW model by considering the global geometric relationships among Pairs of Identical Words (PIWs). A normalized histogram is created that is based on angles between these identical visual words, which are termed PIWAH (Pairs of Identical Visual Words Angle Histogram). The PIWAH representation is invariant to geometrical transformations i.e. scaling and translation but is sensitive to rotation [16, 3]. Anwar *et al.* [3] extend their work to acquire rotation-invariance, and propose Triplets of Identical Visual Words Angle Histogram (TIWAH), by computing relationship between triplets of identical visual words (TIWs). However, the approaches based on computing the relationships between co-occurrences of visual words are known to be computationally expensive.

The second group encompasses methods that divide an image into sub-regions of different shapes; the information about visual words are computed from each of the selected regions. The most notable work in this domain is that of Lazebnik *et al.* [18], who originally proposed SPM. In SPM, an image is divided into rectangular subregions and visual word statistics are aggregated from each region. The final histogram is the concatenation of histograms extracted from each region. Anwar *et al.* [2] present a dense sampling based BoVWs approach and evaluate three types of schemes, i.e. rectangular tiling, log-polar tiling and circular tiling, for culture heritage classification-based problems.

To prove the effectiveness of the proposed approach, in addition to looking at methods concurrent to our approach [18, 37, 15], we have also selected some recent state-of-the-art papers which are focussed on different approaches such as feature extraction [39, 8, 9] and intermediate feature representation [21, 22, 7] techniques to improve classification performance. Mekhalfi *et al.* [21] propose a novel scheme to compactly represent images, using a compressive sensing and multi-feature framework. Their method result in substantial performance gain in comparison to the state-of-the-art methods on land-use image dataset. To attain better classification performance, Zou *et al.* [39] propose the creation of a fusion of local and global image features. For this they first extract local features by using BoVW and SPM. In order to extract global features, they employ multi-scale CLBP (MS-CLBP). For feature representation, they use Kernel Collaborative Representation-based Classification (KCRC). After the representation, residuals are obtained from the two types of features, and the label is assigned based on the sum of the weighted residuals.

Zhu *et al.* [38] propose a fusion strategy of local and global features for high spatial resolution (HSR) scene image classification. In their paper [9], Bian *et al.* propose the fusion of local and global descriptors to enhance the classification performance. They enrich the feature representations by combining both global structures and fine local details of the image scene. To extract global rotation-invariant features; a global saliency-based multiscale, multiresolution and multistructure LBP is employed, and the local Codebookless Model (CLM) is used to represent local discriminative features. Bian *et al.* report better or comparable performance to the state-of-the-art classification methods. Chen *et al.* [10] propose a texture based paradigm, i.e. a Multi-Scale Completed local Binary Patterns (MS-CLBP) descriptor, for land use scene classification. CLBP is known to be effective for rotation-invariant texture classification. They enhance the CLBP to characterize the dominant texture features in multiple resolutions and show consistent increase in performance compared to the state-of-the-art.

Object Bank (OB) is a high-level image representation that encodes the spatial and semantic information [19]. However OB approach suffers from drawback of high-dimensionality and various approaches have been proposed in literature to reduce the dimensions and enhance the performance of OB [19, 34]. To boost the performance of OB representation Zang *et al.* [34] proposed a threshold value filter method. They used Matthew effect normalization method to simplify OB representation and constructed more compact descriptors and showed improved performance on three real-world datasets, with substantial dimensionality reduction of image descriptors.

Cheng *et al.*[11] propose the Bag of Convolutional Features (BoCF) that uses deep convolutional features to generate a visual vocabulary for remote sensing image classification. Lu *et al.* [14] propose an unsupervised representation learning method for remote sensing scene classification. Scott *et al.* [27] propose to use Transfer Learning (TL) in combination with fine-tuning and augmentation. They evaluate the effectiveness of their proposed approach on high-resolution imagery classification and achieve significantly higher accuracy than the most effective methods. Scott *et al.* [28] investigate a variety of fusion techniques including voting, weighted averages, and fuzzy integrals, to blend multiple DCNN land cover classifiers; as CaffeNet, GoogLeNet, and ResNet50, into a single aggregate classifier. It is worth mentioning here that deep learning and neural network based approaches require huge amounts of data (in the millions) and time for training [17]. The BoVW model is a plug-n-play method which can be adopted without any prior initialization or training.

3. Proposed Methodology

This section provides an overview of the BoVW model along with its basic notations as well as an explanation of the proposed approach based on histograms of concentric weighted circles.

The BoVW Model

The BoVW model is based on text-based retrieval system used extensively in the literature to compare documents [29]. BoVW represents images by using histograms of visual words. Visual words are analogues to words in a document, and represent small picture elements (local areas in images). BoVW image histogram representation is order-less and is created by ignoring the locations of visual words in the 2-D image. Histogram intersection is used to determine the similarity of images.

In BoVW, as a first step, each image *I* is represented by a set of image descriptors as in (1)

$$Im = \{d_1, d_2, d_3, ..., d_I\} \tag{1}$$

where d_i is the color, shape or texture description of the patch of an image, and *I* denotes the total number of image descriptors.

To create a visual vocabulary, an unsupervised clustering technique *K*-means is applied on the extracted descriptors (2)

$$v = \{w_1, w_2, w_3, ..., w_K\} \tag{2}$$

where *K* is the predefined number of clusters or visual words.
The mapping of a visual word to the nearest descriptor is performed according to (3)

$$w(d_j) = \underset{w \in v}{argmin} Dist(w, d_j) \tag{3}$$

Here, w(d_j) depicts the visual word assigned to the j^{th} descriptor and *Dist* (w,d_j) signifies the distance between the descriptor d_j and the visual word *w*.
Clustering reduces the high dimensional feature space; hence as a result of clustering, each descriptor is mapped to a visual word and the final representation of the image is the histogram of visual words. The count of bins in the histogram equals the number of visual words in the dictionary (i.e. *K*). If each bin represents a visual word w_i in *voc*, (4)

$$bin_i = card(D_i) \;\; where \;\; D_i = \{d_j, j \in 1, ..., n \mid w(d_j) = w_i\} \tag{4}$$

D_i is the set of all the descriptors that belong to a particular visual word w_i in an image. The cardinality of set D_i is given by $Card(D_i)$, which gives the count of the elements of the set. This is repeated for every word in the image to obtain the final representation. The histogram thus created does not retain the spatial information of the interest points.

Concentric Weighted Circles Histogram (CWCH)

To add the spatial contents of the image and achieve rotation-invariance, we propose creating a histogram of concentric circles. A weighted scheme is applied to represent the image in the form of a histogram of visual words. The image is partitioned into segments at different levels in a concentric circles fashion, where the k^{th} level has $k+1$ number of segments. Each extracted segment is represented by a histogram of visual words.

For an image *I* of size $R \times C$, the centroid $c = (c_x, c_y)$ of an image is calculated as

$$c_x = \frac{1}{\mid I \mid} \sum_{i=1}^{|I|} x_i, \;\; c_y = \frac{1}{\mid I \mid} \sum_{i=1}^{|I|} y_i \tag{5}$$

where $I = \{(x_i, y_i) \mid 1 \leq x_i \leq C, 1 \leq y_i \leq R\}$ and $\mid I \mid$ is the number of elements in *I*. Let L be the number of levels, then the radius r of k^{th} level is given by

$$r_k = \frac{k}{L} \, min\{c_x, c_y\} \tag{6}$$

The radius of the smallest circle will be r_1. The similarity of two images with *L* number of levels is determined as

$$S = \sum_{k=0}^{L} w_k N_k \tag{7}$$

where $w_k = 1/2^{L-k}$ is the weight assigned to each level and N_k is the difference of histogram intersection of two consecutive levels.
The histogram intersection of two images is calculated as

$$HI(P, Q) = \sum_{i=1}^{c} min(P_i, Q_i) \tag{8}$$

where *P* and *Q* are histograms with *c* bins, and P_i denotes the count of i_{th} bin of *P*.

The block diagram of the proposed research is shown in Fig. 2. There are three types of commonly used feature-sampling techniques i.e. dense sampling, dense interest points and interest points. We have selected dense interest points, as it is a hybrid approach and combines the best of both worlds i.e. interest points and dense sampling [31]. For feature extraction, all of the images are converted to gray scale, and dense SIFT with multi-scale and steps-size 8, is used to create feature descriptors. These descriptors are clustered into similar groups using the *k*-means clustering technique to create the visual vocabulary. Due to the unsupervised nature of *k*-means clustering, the experiments are repeated 10 times, with randomly selected subsets of training and test images. The average values are reported in tables and graphs. The size of visual vocabulary is a major parameter that has a significant impact on the performance of system. The performance is directly proportional to vocabulary size, and a large vocabulary size tends to overfit [23]. To determine the optimal performance of the proposed image representations, experiments are conducted with visual vocabularies of different sizes.

The image space is partitioned into concentric circles, appropriate weights are assigned to each level and the histograms from each level are concatenated to obtain the final representation. The image representations obtained from the training set are then used to train the classifier; evaluation is done with the test sets. For classification, Support Vector Machines (SVM), which is characterized as a supervised learning method [35], is used. SVM acquires the ability to generate non-linear decision boundaries and uses the kernel method to compute the dot product in a high dimensional feature space. Given positive and negative training images, the objective is to classify test images into their respective object classes. The histograms constructed by using the proposed framework are normalized and SVM Hellinger Kernel [32] is applied to the normalized histograms. The SVM Hellinger kernel is selected because of its low computational cost. To determine the optimal value for the regularization parameter *C*, 10-fold cross validation is applied on the training dataset.

4. Experiments and Results

To evaluate the effectiveness of the proposed approach, it is applied for natural scene classification using a 15-scene image benchmark, as it is a challenging and the most widely used dataset in the literature so far [39]. Experiments are also conducted for remote sensing image scene classification, using the well-known UC Merced Land Use dataset. For all datasets, we followed the same experimental setup to create the proposed image representations. To conduct experiments for rotation-invariance, a rotation dataset is created from each dataset, consistent with related work [15]. This section provides an overview of the experimental results and of the image benchmarks used to evaluate the proposed research.

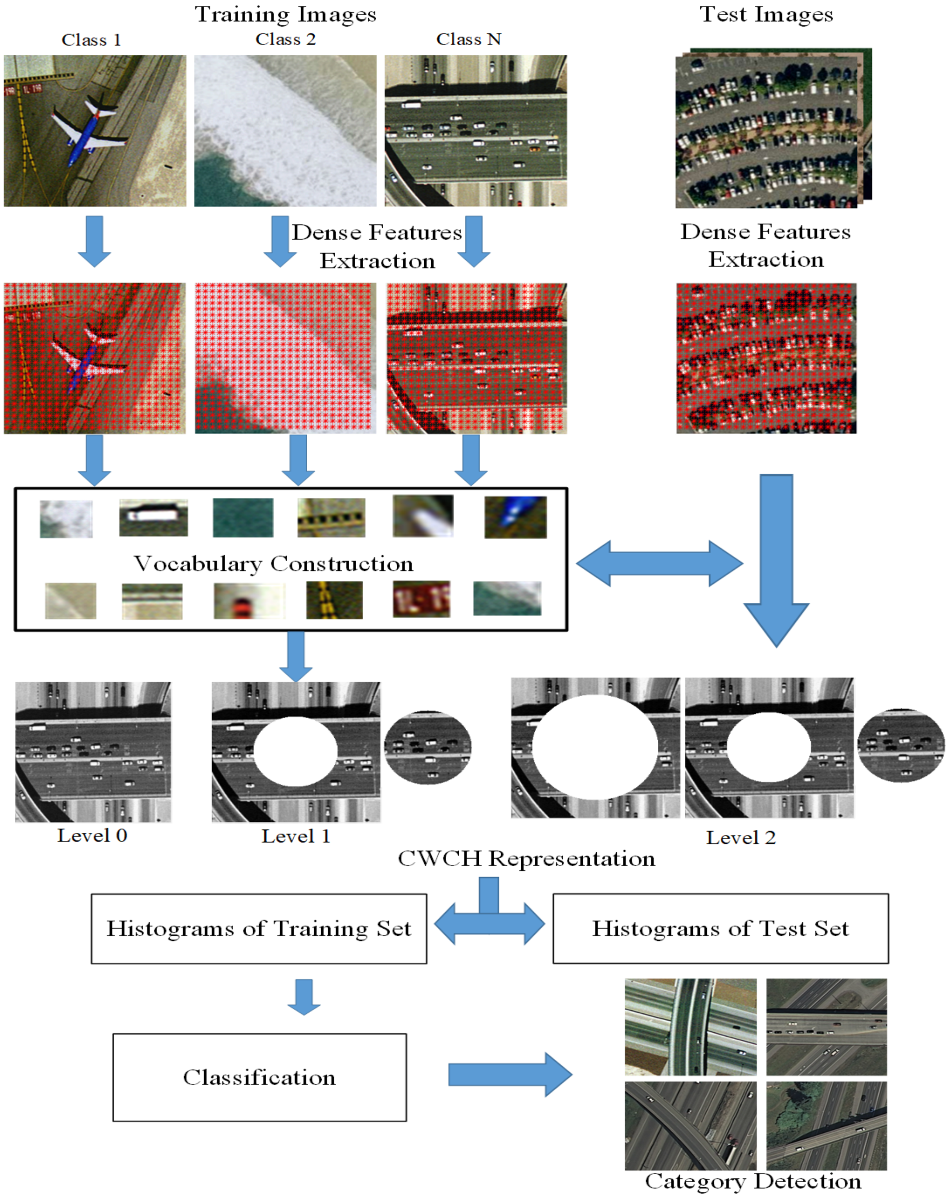


Fig. 2. The block diagram of the proposed framework

15-scene Image Dataset

The first dataset used in our experiments is the well-known 15-scene image dataset comprised of indoor and outdoor scene categories. Fig. 3 shows the examples of images for a 15-scene dataset, indicating names and number of images per category. Initially 8 categories were contributed by Oliva and Torralba [25], Li and perona [13] added 5 classes, and the rest were collected by [18]. The predominant sources of images are the Internet, personal photographs and the COREL collection. It is a challenging dataset, and has a total of 4485 images with an average image size of $300 * 250$ pixels. It is the most widely used dataset so far.

To ensure a fair comparison, the testing and training samples are chosen in accordance with the state-of-the-art methods [18, 15]. The training set is comprised of 100 randomly selected images and the rest of the images are used for testing; hence the total number of training images is 1500 and there are 2985 test images. The class numbers in our experiments are ‘Bedroom = 1’, ‘Calsubrub = 2’, ‘Industrial = 3’, ‘Kitchen= 4’, ‘Living room = 5’, ‘MITcoast = 6’, ‘MITforest = 7’, ‘MIThighway = 8’, ‘MITinsidecity = 9’, ‘MITmountain = 10’, ‘MITopen country = 11’, ‘MITstreet = 12’, ‘MITtallbuilding = 13’, ‘PARoffice = 14’ and ‘Store = 15’.

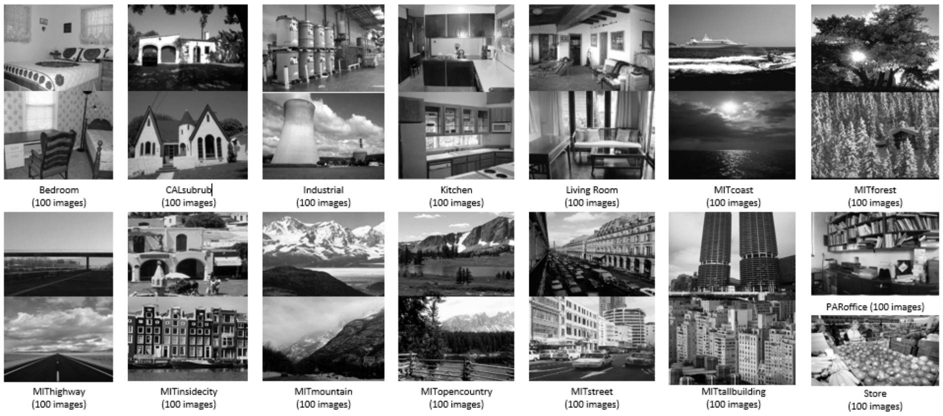


Fig. 3. Example images from each category of the 15-scene image dataset [18]

To obtain the optimal performance for accurate feature representations, experiments are performed with different sizes of visual vocabulary on both the 15-scene and the rotated dataset. Fig. 4 demonstrates the robustness of the proposed approach on two datasets and provides a graphical representation of the performance on vocabulary of different sizes. For the two datasets, optimal performance for CWCH is obtained for a vocabulary with a size of 400. The dimensions of the resultant feature vector are 2400.

Table 1 provides a comparison of our proposed CWCH with the recent methods that incorporate spatial context in the BoVW model. Experimental results demonstrate the robustness of our approach compared with the state-of-the-art methods.

It is evident from the table that our method provides the highest accuracy compared to the state-of-the-art methods. Though the dimensions of PIWAH [16] are low compared to CWCH, our approach is relatively simple and achieves 12.04% higher accuracy compared to their method. The dimensions of CWCH can be reduced further by applying some dimension reduction technique without significant loss in accuracy. Our method achieves

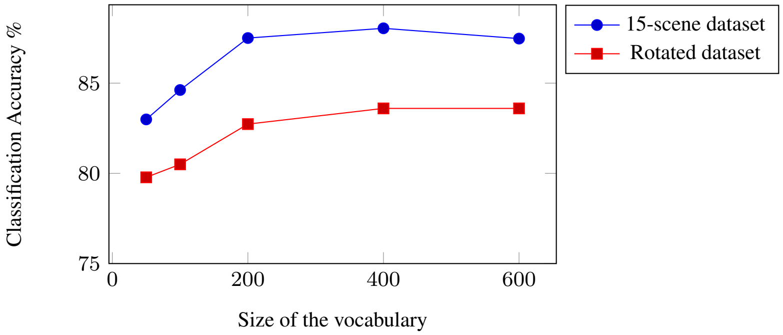


Fig. 4. Mean average accuracy as a function of vocabulary size using 15-scene image dataset

Table 1. Classification accuracy comparison of the proposed CWCH with the state-of-the-art methods

Algorithms	Feature Dimensionality	Accuracy
PIWAH [16]	1800	76%
SPM [18]	4200	81.1%
Zang *et al.* [34]	3717	81.5%
$\text{SPS}_{\text{ad}}+$ [17]	13200	83.7%
Karmakar *et al.* [15]	4200	84.20%
LGF [39]	X	85.8%
CWCH	2400	**88.04%**

6.94% higer accuracy compared to SPM [18], which is the pioneering work to incorporate spatial information into the BoVW model.

In another recent paper [34], Zang *et al.* reduce the dimensions of Object Bank (OB) achieving an accuracy 81.5%, CWCH produced a low-dimensional histogram representation with a 6.54% higher accuracy. Compared with the approach of rectangular rings [15] our method achieves 3.84% more accuracy with fewer dimensions reported in their work. Our method clearly outperforms the state-of-the-art methods in terms of accuracy and dimensions with an accuracy 88.04%.

In order to demonstrate the sustainable performance of the proposed image representation, we have performed a class-wise comparison with the state-of-the-art method [39] shown in Fig. 5. Zou *et al.*[39] consider the spatial context by incorporating SPM [18] into their implementation.

The average confusion matrix for 15-scene image dataset is shown in Fig. 6. The diagonal values show the precision normalized percentages of each class.

UC Merced Land-Use (UCM) Image Dataset

The UCM image dataset was created by Yang and Newsam [33] and is comprised of 21 land-use image scene categories. This dataset has a large geographical scale and the images have been downloaded from the United States Geological Survey (USGS) National map. Each class consists of 100 images with an average size of 256×256 pixels. Fig. 7

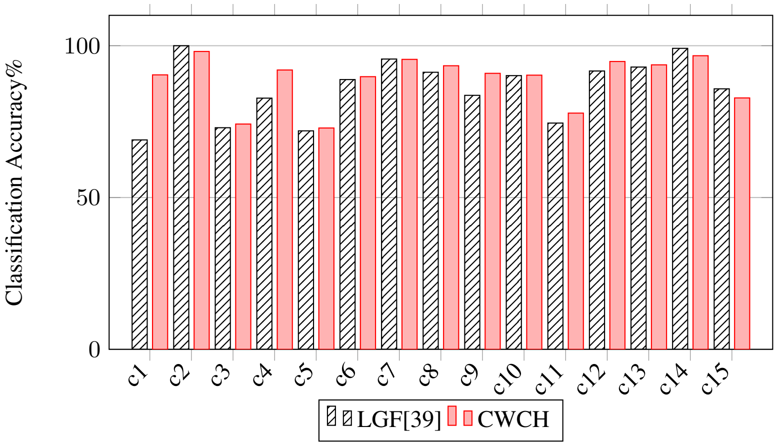
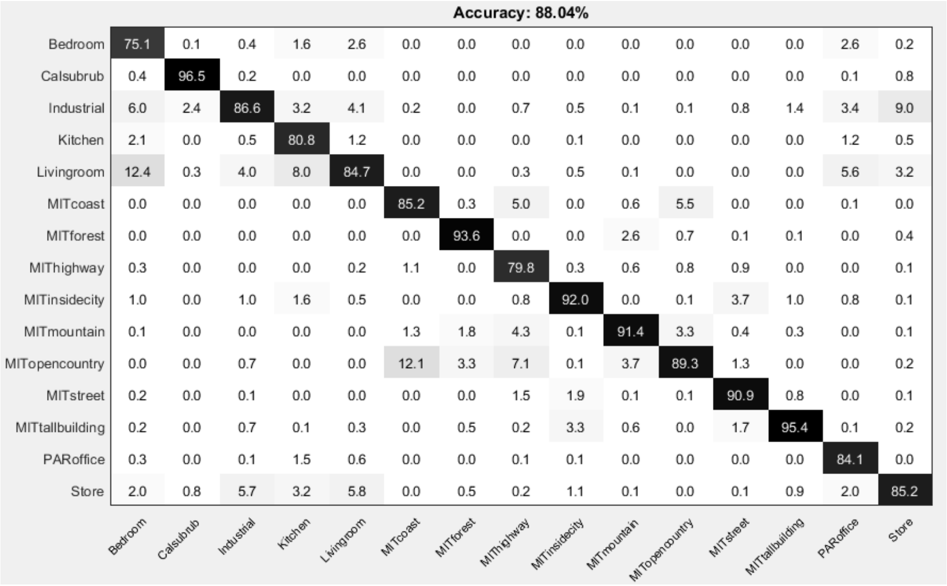


Fig. 5. Class-wise comparison between LGF [39] and CWCH for 15-scene image dataset



Accuracy: 88.04%

	Bedroom	Calsubrub	Industrial	Kitchen	Livingroom	MITcoast	MITforest	MIThighway	MITinsidecity	MITmountain	MITopencountry	MITstreet	MITtallbuilding	PARoffice	Store
Bedroom	75.1	0.1	0.4	1.6	2.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.6	0.2
Calsubrub	0.4	96.5	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.8
Industrial	6.0	2.4	86.6	3.2	4.1	0.2	0.0	0.7	0.5	0.1	0.1	0.8	1.4	3.4	9.0
Kitchen	2.1	0.0	0.5	80.8	1.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	1.2	0.5
Livingroom	12.4	0.3	4.0	8.0	84.7	0.0	0.0	0.3	0.5	0.1	0.0	0.0	0.0	5.6	3.2
MITcoast	0.0	0.0	0.0	0.0	0.0	85.2	0.3	5.0	0.0	0.6	5.5	0.0	0.0	0.1	0.0
MITforest	0.0	0.0	0.0	0.0	0.0	0.0	93.6	0.0	0.0	2.6	0.7	0.1	0.1	0.0	0.4
MIThighway	0.3	0.0	0.0	0.0	0.2	1.1	0.0	79.8	0.3	0.6	0.8	0.9	0.0	0.0	0.1
MITinsidecity	1.0	0.0	1.0	1.6	0.5	0.0	0.0	0.8	92.0	0.0	0.1	3.7	1.0	0.8	0.1
MITmountain	0.1	0.0	0.0	0.0	0.0	1.3	1.8	4.3	0.1	91.4	3.3	0.4	0.3	0.0	0.1
MITopencountry	0.0	0.0	0.7	0.0	0.0	12.1	3.3	7.1	0.1	3.7	89.3	1.3	0.0	0.0	0.2
MITstreet	0.2	0.0	0.1	0.0	0.0	0.0	0.0	1.5	1.9	0.1	0.1	90.9	0.8	0.0	0.1
MITtallbuilding	0.2	0.0	0.7	0.1	0.3	0.0	0.5	0.2	3.3	0.6	0.0	1.7	95.4	0.1	0.2
PARoffice	0.3	0.0	0.1	1.5	0.6	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	84.1	0.0
Store	2.0	0.8	5.7	3.2	5.8	0.0	0.5	0.2	1.1	0.1	0.0	0.1	0.9	2.0	85.2

Fig. 6. Confusion matrix for 15-scene image dataset

shows the example images from each class along with the class label and the total number of images in each category.

The class numbers in our experiments are ‘Agricultural=1’, ‘Airplane=2’, ‘Baseball diamond=3’, ‘Beach=4’, ‘Buildings=5’, ‘Chaparral=6’, ‘Dense residential=7’, ‘Forest=8’, ‘Freeway=9’, ‘Golf Course=10’, ‘Harbor=11’, ‘Intersection=12’, ‘Medium residential=13’, ‘Mobile home park=14’, ‘Overpass=15’, ‘Parking lot=16’, ‘River=17’, ‘Runway=18’, ‘Sparse residential=19’, ‘Storage tanks=20’, ‘Tennis court=21’. The training and the test

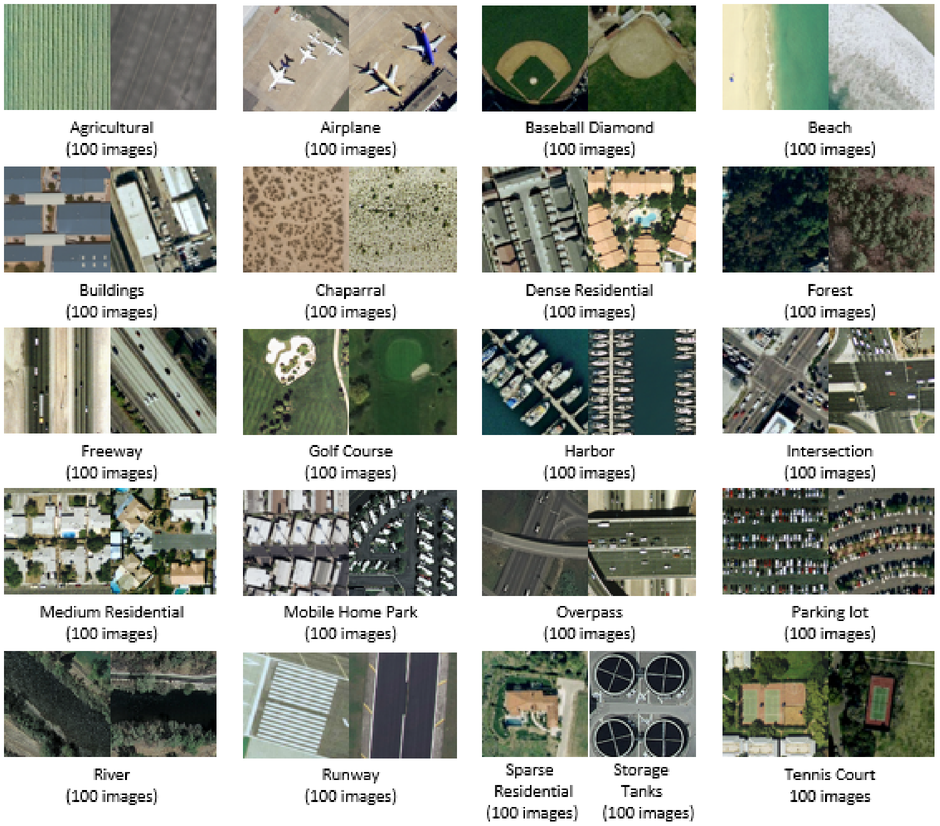


Fig. 7. Example images from each category of the UCM image dataset [33]

set are chosen in accordance with related works [39, 9, 33] for the sake of comparison. For training, 80 images are randomly chosen from each class and the remaining are used for testing.

The performance comparison of proposed approach for UCM and the rotated dataset over vocabulary of different sizes is shown in Fig. 8. It can be seen that the proposed approach is robust to image rotations. Using the proposed approach on both daasets, we obtain optimal performance for a vocabulary with a size of 200, resulting in a 1200 dimensional feature vector.

In table 2 we present a comparison of our proposed CWCH with the state-of-the-art research. It is evident that CWCH clearly outperforms not only those methods that are consistent with our approach, but also other works that enhance feature representation in fusion with spatial context as well as those that enhance intermediate representations.

In [37], Zhao *et al.* used concentric circle-structured multiscale BoVW model to incorporate the spatial context. While they use multiple features i.e. SIFT, color moments and LBP to enhance their representation, our method provides 12.76% more accuracy compared to their approach. To the best of our knowledge, Scott *et al.* [28] provide the best classification for the UCM dataset. The CWCH representation provides competitive

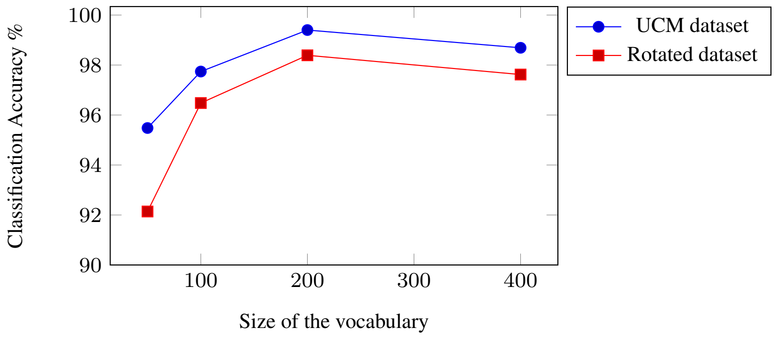


Fig. 8. Mean average accuracy as a function of vocabulary size using the UCM image dataset

Table 2. Classification accuracy comparison of the proposed CWCH with the state-of-the-art methods

Algorithms	Accuracy
SPM [38]	82.3%
BoCF [11]	84.32%
CCM-BOVW [37]	86.64%
MSCLBP$_1$ [10]	90.6%
SOS [21]	94.33%
LGF [39]	95.48%
Lu *et al.* [14]	95.71%
salM^{3}LBP-CLM [9]	95.75%
LGFBOVW [38]	96.88%
ResNet50 [27]	98.5%
Evolved Sugeno [28]	99.33%
CWCH	99.4%

performance to the state-of-the-art, based on feature extraction [39], spatial context [37], the application of intermediate feature representation [21] and deep learning techniques [28].

The class-wise comparison of CWCH with CCM-BOVW [37] is shown in Fig. 9. In CCM-BOVW [37] the spatial information is incorporated by using a concentric-circle based approach, their approach appear good only for the classes that are sensitive to orientations, as airplane, baseball diamond, golf course and storage tanks. Whereas, CCM-BOVW [37] model did not have a significant impact on categories, that have simple pattern and do not suffer from orientations as forest, river, agricultural and chapparal. Our proposed approach shows significant improvement in all the classes.

The confusion matrix for the UCM image dataset is shown in Fig. 10. The diagonal values show the precision normalized percentages of each class.

Complexity Performance

The algorithms are computed using Intel(R) Core i7 (sixth generation) 2.59 GHz CPU, 8 GB RAM and the Windows-10 operating system. The proposed algorithms are imple-

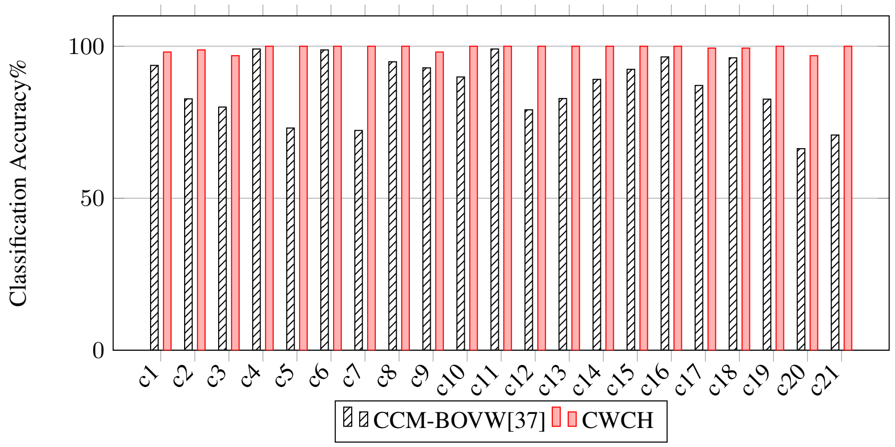
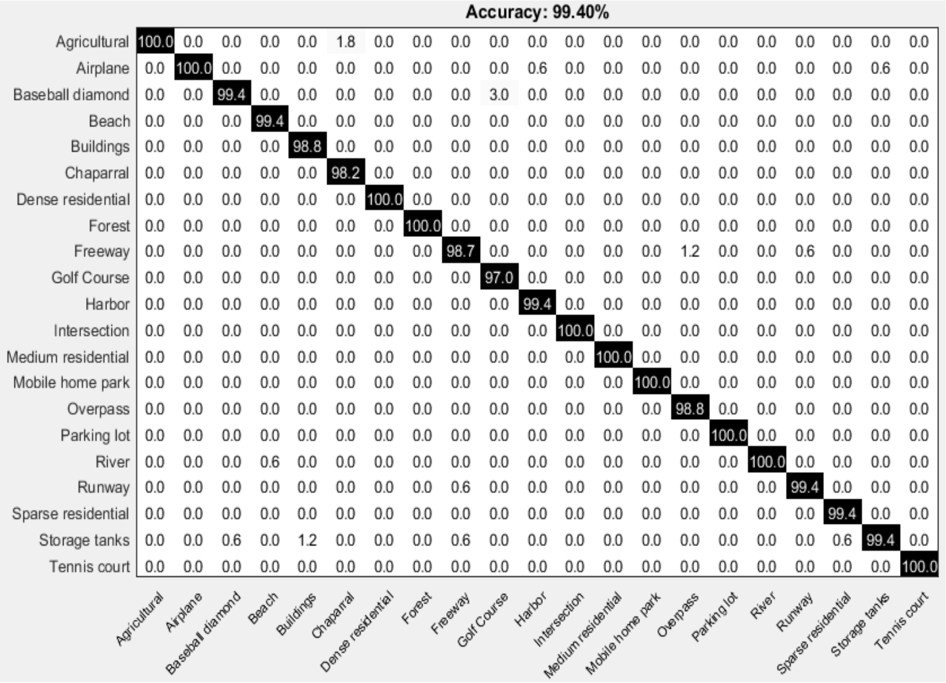


Fig. 9. Class-wise comparison between CCM-BOVW [37] and CWCH for UCM image dataset


Accuracy: 99.40%

	Agricultural	Airplane	Baseball diamond	Beach	Buildings	Chaparral	Dense residential	Forest	Freeway	Golf Course	Harbor	Intersection	Medium residential	Mobile home park	Overpass	Parking lot	River	Runway	Sparse residential	Storage tanks	Tennis court
Agricultural	100.0	0.0	0.0	0.0	0.0	1.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Airplane	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0
Baseball diamond	0.0	0.0	99.4	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Beach	0.0	0.0	0.0	99.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Buildings	0.0	0.0	0.0	0.0	98.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Chaparral	0.0	0.0	0.0	0.0	0.0	98.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Dense residential	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Forest	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Freeway	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.7	0.0	0.0	0.0	0.0	0.0	1.2	0.0	0.0	0.6	0.0	0.0	0.0
Golf Course	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	97.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Harbor	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Intersection	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Medium residential	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mobile home park	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Overpass	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.8	0.0	0.0	0.0	0.0	0.0	0.0
Parking lot	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
River	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Runway	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.4	0.0	0.0	0.0
Sparse residential	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.4	0.0	0.0
Storage tanks	0.0	0.0	0.6	0.0	1.2	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	99.4	0.0
Tennis court	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0

Fig. 10. Confusion matrix for the UCM image dataset

mented in MATLAB and the visual vocabulary (codebook) is constructed offline using a training dataset and tested at run-time using a test dataset. The average CPU time (in seconds) required from features extraction to image classification for CWCH and BoVW on test dataset is presented in Table 3.

Table 3. Classification accuracy and time comparison between BoVW and CWCH

Dataset	BoVW		CWCH	
	Accuracy	Time	Accuracy	Time
15-scene	84.11%	74.40s	88.04%	118.67s
UCM	97.62%	9.93s	99.4%	21.09s

Our method provides better classification rates on both datasets compared to BoVW without spatial information. However, BoVW seems to be computationally efficient, as it does not involve any spatial information extraction step. Comparatively, our approach provides 3.93% more accuracy on a 15-scene image dataset and 1.78% on the UCM image dataset. The computation time of CCM-BOVW [37] is not reported, but it involves multiple feature extraction steps and it also creates multiple multi-resolution images as part of feature extraction. Our approach is simple, as it is comprised of a single feature extraction step and provides better classification accuracy.

5. Conclusion and Future Directions

Due to recent growth in technology, the complexity and volume of multimedia contents is significantly increasing. This has creased a demand for a highly effective image classification system. In this paper, we addressed two problems that are associated with the classification-based framework of the BoVW model. The standard BoVW model lacks spatial information and the approaches based on the division of images into cells to create histograms of visual words do not allow rotations and changes in view point. The proposed approach constructs the histograms of visual words by using concentric circles to handle the changes in view point, rotations and computation of spatial information. A weighted scheme is applied for image representation and performance is evaluated by using two levels of concentric circles. The application of a weighted scheme makes it possible to represent the image in a lower dimensional space that address the problem of over-fitting.

The proposed histogram representation is simple and robust and can handle rotations and changes in view point. Two standard image benchmarks are used for the performance evaluation of the proposed approach. We would like to extend this work by creating a unified approach that is tolerant to other transformations as well, i.e. scaling and translation. The proposed framework can be applied for image classification of a huge volume of multimedia contents. In future, we intend to evaluate our proposed method on larger datasets like ImageNet and Flicker by using deep convolution neural networks. A direct extension of this work could be to incorporate other spatial cues such as color and shape, or to combine it with some complementary relative spatial feature extraction method to enhance the classification performance.

Conflict of Interest. Authors Bushra Zafar, Rehan Ashraf, Nouman Ali, Mudassar Ahmed, Sohail Jabbar, Kashif Naseer, Awais Ahmad, and Gwanggil Jeon declare that they have no conflict of interest.

Acknowledgments. This work was supported by National Natural Science Foundation of China (NSFC) under Grant 61771378.

References

1. Ali, N., Bajwa, K.B., Sablatnig, R., Mehmood, Z.: Image retrieval by addition of spatial information based on histograms of triangular regions. Computers & Electrical Engineering 54, 539–550 (2016)
2. Anwar, H., Zambanini, S., Kampel, M.: Supporting ancient coin classification by image-based reverse side symbol recognition. In: International Conference on Computer Analysis of Images and Patterns. pp. 17–25. Springer (2013)
3. Anwar, H., Zambanini, S., Kampel, M.: Encoding spatial arrangements of visual words for rotation-invariant image classification. In: German Conference on Pattern Recognition. pp. 443–452. Springer (2014)
4. Anwar, H., Zambanini, S., Kampel, M.: Efficient scale-and rotation-invariant encoding of visual words for image classification. IEEE Signal Processing Letters 22(10), 1762–1765 (2015)
5. Ashraf, R., Ahmed, M., Ahmad, U., Habib, M.A., Jabbar, S., Naseer, K.: Mdcbir-mf: multimedia data for content-based image retrieval by using multiple features. Multimedia Tools and Applications pp. 1–27 (2018)
6. Ashraf, R., Ahmed, M., Jabbar, S., Khalid, S., Ahmad, A., Din, S., Jeon, G.: Content based image retrieval by using color descriptor and discrete wavelet transform. Journal of medical systems 42(3), 44 (2018)
7. Ashraf, R., Bajwa, K.B., Mahmood, T.: Content-based image retrieval by exploring bandletized regions through support vector machines. J. Inf. Sci. Eng. 32(2), 245–269 (2016)
8. Ashraf, R., Bashir, K., Irtaza, A., Mahmood, M.T.: Content based image retrieval using embedded neural networks with bandletized regions. Entropy 17(6), 3552–3580 (2015)
9. Bian, X., Chen, C., Tian, L., Du, Q.: Fusing local and global features for high-resolution scene classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2017)
10. Chen, C., Zhang, B., Su, H., Li, W., Wang, L.: Land-use scene classification using multi-scale completed local binary patterns. Signal, image and video processing 10(4), 745–752 (2016)
11. Cheng, G., Li, Z., Yao, X., Guo, L., Wei, Z.: Remote sensing image scene classification using bag of convolutional features. IEEE Geoscience and Remote Sensing Letters 14(10), 1735–1739 (2017)
12. Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Workshop on statistical learning in computer vision, ECCV. pp. 1–2. Prague (2004)
13. Fei-Fei, L., Perona, P.: A bayesian hierarchical model for learning natural scene categories. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. vol. 2, pp. 524–531. IEEE (2005)
14. Fu, M., Yuan, Y., Lu, X.: Remote sensing scene classification by unsupervised representation learning. IEEE Trans. Geoscience and Remote Sensing 55(9), 5148–5157 (2017)
15. Karmakar, P., Teng, S.W., Lu, G., Zhang, D.: Rotation invariant spatial pyramid matching for image classification. In: Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on. pp. 1–8. IEEE (2015)
16. Khan, R., Barat, C., Muselet, D., Ducottet, C.: Spatial orientations of visual word pairs to improve bag-of-visual-words model. In: Proceedings of the British Machine Vision Conference. pp. 89–1. BMVA Press (2012)
17. Khan, R., Barat, C., Muselet, D., Ducottet, C.: Spatial histograms of soft pairwise similar patches to improve the bag-of-visual-words model. Computer Vision and Image Understanding 132, 102–112 (2015)

18. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Computer vision and pattern recognition, 2006 IEEE computer society conference on. vol. 2, pp. 2169–2178. IEEE (2006)
19. Li, L.J., Su, H., Lim, Y., Fei-Fei, L.: Object bank: An object-level image representation for high-level visual recognition. International journal of computer vision 107(1), 20–39 (2014)
20. Ma, L., Li, M., Ma, X., Cheng, L., Du, P., Liu, Y.: A review of supervised object-based land-cover image classification. ISPRS Journal of Photogrammetry and Remote Sensing 130, 277–293 (2017)
21. Mekhalfi, M.L., Melgani, F., Bazi, Y., Alajlan, N.: Land-use classification with compressive sensing multifeature fusion. IEEE Geoscience and Remote Sensing Letters 12(10), 2155–2159 (2015)
22. Nazir, A., Ashraf, R., Hamdani, T., Ali, N.: Content based image retrieval system by using hsv color histogram, discrete wavelet transform and edge histogram descriptor. In: Computing, Mathematics and Engineering Technologies (iCoMET), 2018 International Conference on. pp. 1–6. IEEE (2018)
23. Nowak, E., Jurie, F., Triggs, B.: Sampling strategies for bag-of-features image classification. Computer Vision–ECCV 2006 pp. 490–503 (2006)
24. O'Hara, S., Draper, B.A.: Introduction to the bag of features paradigm for image classification and retrieval. arXiv preprint arXiv:1101.3354 (2011)
25. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision 42(3), 145–175 (2001)
26. Savarese, S., Winn, J., Criminisi, A.: Discriminative object class models of appearance and shape by correlatons. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. vol. 2, pp. 2033–2040. IEEE (2006)
27. Scott, G.J., England, M.R., Starms, W.A., Marcum, R.A., Davis, C.H.: Training deep convolutional neural networks for land–cover classification of high-resolution imagery. IEEE Geoscience and Remote Sensing Letters 14(4), 549–553 (2017)
28. Scott, G.J., Marcum, R.A., Davis, C.H., Nivin, T.W.: Fusion of deep convolutional neural networks for land cover classification of high-resolution imagery. IEEE Geoscience and Remote Sensing Letters 14(9), 1638–1642 (2017)
29. Sivic, J., Zisserman, A.: Video google: A text retrieval approach to object matching in videos. In: null. p. 1470. IEEE (2003)
30. Song, Y., McLoughlin, I.V., Dai, L.R.: Local coding based matching kernel method for image classification. PloS one 9(8), e103575 (2014)
31. Tuytelaars, T.: Dense interest points. In: Computer Vision and Pattern Recognition. IEEE (2010)
32. Vedaldi, A., Zisserman, A.: Sparse kernel approximations for efficient classification and detection. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. pp. 2320–2327. IEEE (2012)
33. Yang, Y., Newsam, S.: Bag-of-visual-words and spatial extensions for land-use classification. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. pp. 270–279. ACM (2010)
34. Zang, M., Wen, D., Liu, T., Zou, H., Liu, C.: A pooled object bank descriptor for image scene classification. Expert Systems with Applications 94, 250–264 (2018)
35. Zhang, D., Islam, M.M., Lu, G.: A review on automatic image annotation techniques. Pattern Recognition 45(1), 346–362 (2012)
36. Zhang, J., Marszałek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: A comprehensive study. International journal of computer vision 73(2), 213–238 (2007)
37. Zhao, L.J., Tang, P., Huo, L.Z.: Land-use scene classification using a concentric circle-structured multiscale bag-of-visual-words model. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(12), 4620–4631 (2014)

38. Zhu, Q., Zhong, Y., Zhao, B., Xia, G.S., Zhang, L.: Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery. IEEE Geoscience and Remote Sensing Letters 13(6), 747–751 (2016)
39. Zou, J., Li, W., Chen, C., Du, Q.: Scene classification using local and global features with collaborative representation fusion. Information Sciences 348, 209–226 (2016)

Bushra Zafar is serving as lecturer, department of computer science at Government College University Faisalabad (GCUF), Pakistan. Currently, she is pursuing the Ph.D. degree in computer science with National Textile University (NTU), Faisalabad, Pakistan. Her research interests include Computer Vision, Digital Image processing and Complex Adaptive Systems.

Rehan Ashraf (corresponding author) is currently serving as Assistant Professor at the Department of Computer Science, National Textile University, Faisalabad, Pakistan. He completed his MS Computer Engineering in 2011 from Centre for Advanced Studies in Engineering (CASE), Islamabad, Pakistan and Ph.D in Computer Engineering from University of Engineering and Technology, Taxila, Pakistan. He has published number of research papers in Reputed Journals and Conferences. His areas of interests are Content Base Image Retrieval (CBIR), Digital Image Processing, Machine learning techniques and Computer Vision. He also served as a reviewer for reputed journals such as Mathematical Problems in Engineering, Computer Networks, Multimedia Tools and Applications, and Transactions on Internet and Information Systems.

Nouman Ali is working as an Assistant Professor with the Department of Software Engineering at Mirpur University of Science and Technology (MUST), Mirpur, Azad-Kashmir, Pakistan. He received his PhD in Computer Engineering from the University of Engineering and Technology (UET), Taxila, Pakistan. His research interests include Computer Vision and Digital Image Processing. He also served as a reviewer for reputed journals such as Applied Soft Computing, Digital Signal Processing, Public Library of Science and Transactions on Internet and Information Systems.

Mudassar Ahmad is serving as Assistant Professor in Department of Computer Science, National Textile University, Pakistan. He has 17 Years experience as Network Manager in a Textile Industry. He is an Associate Editor in IEEE Newsletters. His research work is published in many conferences and journals. His research includes Internet of Things, Bid Data and Health care.

Sohail Jabbar is Assistant Professor at Department of Computer Science, and Director of Graduate Programs at Faculty of Sciences, National Textile University, Faisalabad Pakistan. He was Post-Doctoral Researcher at Kyungpook National University, Daegu, South Korea. He has been engaged in many National and International Level Projects. He has authored 1 Book, 2 Book Chapters and 60+ research papers. He is currently engaged as TPC member/chair in many conferences. He is guest editor of Sis in various Journals of Elsevier, Springer, KIPS and Taylor & Francis. Sohail is on collaborative research with renowned research centers and institutes around the globe on various issues in the domains of Internet of Things, Wireless Sensor Networks and Big Data.

Muhammad Kashif Naseer is a Junior Lecturer in Department of Computer Engineering at Bahria University Islamabad, Pakistan. He is pursuing Ph.D. in Electrical Engineering from the same University. His research interests include 5G, IoT, AI, Big Data, Cloud computing, Networking and Embedded Systems. He received the BS, MS degrees from Air University and UET Peshawar in 2007 and 2014 respectively.

Awaise Ahmad received the B.S. degree (CS) from the University of Peshawar, Peshawar, Pakistan, and the M.S. degree (telecommunication and networking) from Bahria University, Islamabad, Pakistan, in 2008 and 2010, respectively. Currently, he is working with Department of Computer Science, Bahria University Islamabad Pakistan. Previsouly, he was Assistant Professor at Yeungnam University, South Korea. His current research work includes data science, Big Data analytics, machine-to-machine communication, IoT, and wireless sensor network. Dr. Ahmad was the recipient of three prestigious awards: 1) Research Award from President of Bahria University Islamabad, Pakistan in 2011, 2) Best Paper Nomination Award in WCECS 2011 at UCLA, USA, and 3) Best Paper Award in 1st Symposium on CSE, Moju Resort, Korea, in 2013.

Gwanggil Jeon (corresponding author) received the B.S., M.S., and Ph.D. (summa cum laude) degrees from the Department of Electronics and Computer Engineering, Hanyang University, Seoul, Korea, in 2003, 2005, and 2008, respectively. He has been with Hanyang University, University of Ottawa, Niigata University, Incheon National University, and Xidian University. His current research interests include image processing, particularly image compression, motion estimation, demosaicking, and image enhancement, and computational intelligence, such as fuzzy and rough sets theories. Dr. Jeon was a recipient of the IEEE Chester Sall Award in 2007 and the ETRI Journal Paper Award in 2008.

Received: January 5, 2018; Accepted: August 29, 2018.