ICIC Express Letters
Volume **13**, Number **2**, February **2019**
 ISSN 1881-803X
pp. **119–126**

EDGE-SIFT FOR IMPROVED IMAGE CLASSIFICATION USING MULTI-SCALE GABOR

Weihan Wang, Ying Zhou and Xiaoyan Yu*

School of Physics and Electronic Engineering
Harbin Normal University
No. 1, Shida Road, Linmin Economic Development Zone, Harbin 150025, P. R. China
*Corresponding author: 939700576@qq.com

Received August 2018; accepted November 2018

Abstract. *BOW (Bag-of-Word) model is one of the most popular methods for image classification which represents an image as an orderless collection of local descriptors. Although the BOW is an oversimplified and effective method, it discards the spatial information and induces the redundant information of descriptors. To overcome these weaknesses, we propose an improved algorithm with edge-SIFT (Edge Scale-Invariant Feature Transform) and SPM (Spatial Pyramid Matching) method. First, edge-SIFT descriptors are extracted from an edge image obtained after applying multi-scale Gabor filters to a color image. Next, spatial visual histograms are constructed by using SPM model. Finally, image classification is realized by utilizing SVM (Support Vector Machine) to train and test three benchmark datasets: Caltech101, Caltech256 and CompCars to conduct a quantitate evaluation of the proposed algorithm. Experimental results show that the proposed approach outperforms the similar algorithms.*
Keywords: Edge-SIFT, Gabor, SPM, Image classification

1. **Introduction.** Image classification is able to automatically assign an unknown image to a category according to its visual contents, which has been one of popular research directions in pattern recognition and computer vision [1]. Automatic image classification is capable of quick and efficient query and management of large-scale image databases. It not only saves labor costs, but also ensures high classification accuracy [2].

Feature extraction is one of the most critical steps in image classification, so that different kinds of descriptors have been widely studied. One of renowned examples is Bag-of-Word (BOW) model proposed by Li et al. [3]. This approach represents an image as a collection of visual words, and the image descriptor is generated based merely on the number of occurrences of some particular visual appearances within the image [4]. However, it suffers from several drawbacks, including the limited semantic description of local descriptors and missing of efficient spatial weights [5]. To overcome these problems, Lazebnik et al. proposed a Spatial Pyramid Matching (SPM) method, which partitions an image into increasingly fine sub-regions and computes histograms of local features extracted from inside each sub-region [6]. The resulting has made a remarkable success on a range of image classification.

The conventional SPM approach constructs a visual histogram by clustering all local features of training images with k-means algorithm [7]. In general, Scale-Invariant Feature Transform (SIFT) [8] descriptor is chosen as local descriptors since it is invariant to image scale, rotation, variation in illumination and moreover it provides robust matching by geometric transformation [9]. However, it is hard to achieve satisfactory results on big datasets with high sample sizes and dimensions [10]. The dense SIFT algorithm has good

DOI: 10.24507/icicel.13.02.119

classification results by obtaining descriptors from each location and utilizing a sampling procedure to reduce the computation cost in SIFT algorithm [11].

The information is relevant when the descriptors are extracted just from the region of interest or foreground. The data extracted from the background do not provide relevant information to the image description, so extracting SIFT descriptors from the whole image may result in suboptimal classification [12]. The Gabor filters, whose kernels are similar to the response of the two-dimensional receptive field profiles of the mammalian simple cortical cell, and exhibit the desirable characteristics of spatial locality, spatial frequency and orientation selectivity [13]. And the multi-scale Gabor filter utilizes the gray change information of the pixel and its surrounding pixels to fuse multi-scale image information. In this paper, the edge-SIFT descriptors are extracted by using dense SIFT descriptors from edge image, which are obtained by multi-scale Gabor filter.

According to the above discussion, an improved algorithm is proposed for image classification incorporating edge-SIFT with SPM. In detail, the proposed algorithm can be delineated like this. First, a color image is converted to edge image by exploiting multi-scale Gabor filters. Second, an edge-SIFT feature matrix is obtained from the edge image. Then, a visual histogram is built by SPM. Finally, the test samples are classified by Support Vector Machine (SVM) [14] classifier. By jointly applying these advanced feature extraction techniques, the proposed algorithm is able to improve the accuracy of image classification considerably on the three different datasets. A detailed explanation is provided for each stage in Section 2 and a number of experimental results are illustrated in Section 3, followed by our conclusion.

2. **Proposed Algorithm.** Edge-SIFT descriptors are proposed by the fact that the discriminative descriptors are able to improve the classification performance. The generation of the proposed approach contains two sequential steps. First, an original image is converted into edge image by multi-scale Gabor filters. Then, the Edge-SIFT descriptors are generated by utilizing dense SIFT algorithm on the edge image. The proposed algorithm makes use of textural and spatial information of edge-SIFT and SPM in order to select optimal interest points for good performance on image classification, which can be partitioned into four parts as shown in Figure 1. In the next subsections, the proposed algorithm is described in more detail.

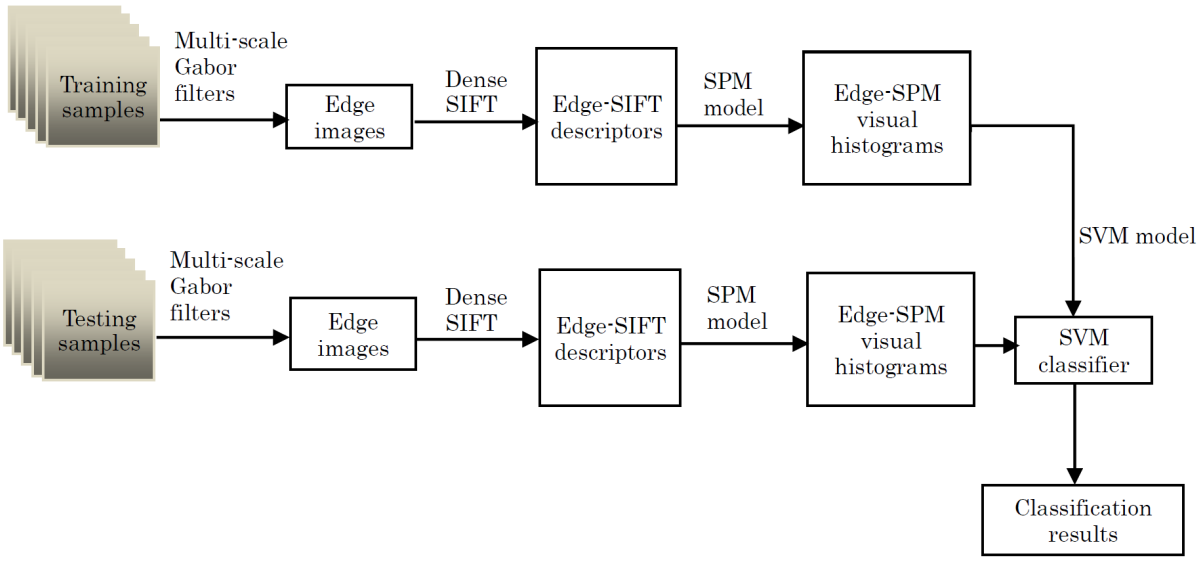


FIGURE 1. Block diagram of the proposed algorithm

2.1. **Multi-scale Gabor filters.** The redundant information is caused by employing SIFT descriptors to extract features from the whole image. Due to the fact that the Gabor filter has scale and orientation tunable property, it is able to extract effective textural image information for classification. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. The imaginary of Gabor filter can be defined as:

$$\varphi(x, y; f, \theta) = \frac{f^2}{\pi\gamma\eta} \exp\left(-\left(\frac{f^2}{\gamma^2}x'^2 + \frac{f^2}{\eta^2}y'^2\right)\right) \sin\left(2\pi f x'\right) \tag{1}$$

where $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$, γ and η are acuteness along the x-axis direction and y-axis direction, f is the central frequency of the filter, and θ is the angle which the modulated plane wave and the Gaussian principal axis rotate counter clockwise. Since a large-scale filter can smooth the image and suppress the noise, it tends to lose detailed grayscale information. Otherwise, small-scale filters are capable of extraction of the detailed grayscale information, while they are sensitive to noises, as a consequence, multi-scale Gabor filters with 3 spatial scales and 16 orientations are employed to extract image edge information.

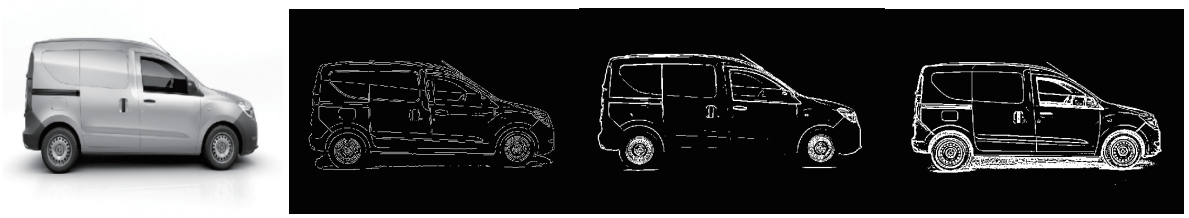
A set of discrete multi-scale Gabor imaginary filters are designed by uniform sampling on direction with $\theta = [0, \pi]$. The proposed filters can be built as:

$$\varphi(m, n; s, k) = \frac{f_s^2}{\pi\gamma\eta} \exp\left(-\left(\frac{f_s^2}{\gamma^2}m'^2 + \frac{f_s^2}{\eta^2}n'^2\right)\right) \sin\left(2\pi f_s x'\right) \tag{2}$$

where $n' = -m\sin\theta_k + n\cos\theta_k$, $m' = m\cos\theta_k + n\sin\theta_k$, $\theta_k = \frac{\pi k}{K}$, $k = 0, 1, \ldots, K-1$, K is the number of samples on direction, θ_k is the angle of the k-th direction, f_s represents the central frequency of $s = \{0, 1, 2\}$ scale. For the input image $I(m, n)$, the imaginary of Gabor filters is obtained on θ_k direction as follows:

$$\zeta(m, n; s, k) = I(m, n) \otimes \varphi(m, n; s, k) = \sum_{m_x}\sum_{n_y} I(m - m_x, n - n_y)\varphi(m, n; s, k) \tag{3}$$

Edge images are extracted by utilizing Canny, Prewitte operators and a multi-scale Gabor algorithm is proposed. The experimental results shown in Figure 2 indicate that the multi-scale Gabor filters achieve the best edge image.



(a) (b) (c) (d)

FIGURE 2. The example of the edge image obtained by Canny, Prewitte and the proposed multi-scale Gabor algorithm: (a) grayscale image, (b) Canny edge image, (c) Prewitte edge image, and (d) multi-scale Gabor edge image

2.2. **Dense SIFT descriptors.** SIFT is a robust descriptor to characterize local gradient information of image pixels. The extended dense SIFT descriptor is used to detected interest points at the dense grids. Owing to omitting the process of constructing Gaussian differential scale space and detecting scale spatial extreme points, the computational complexity is reduced. Dense SIFT has been shown to lead to better performance for various tasks [15]. In this work, edge-SIFT descriptors are extracted from an edge image

by using dense SIFT. The process for extracting edge-SIFT features from an edge image can be divided into three steps. First, the edge image is smoothed by using Gaussian filter. Second, the gradient and angle are calculated from the entire image. Finally, the gradient oriented histograms of 16×16 pixel patches are computed over a grid with spacing of 8 pixels, resulting in a 128-dimensional vector. The goal of our work is to perform image classification based on this edge-SIFT descriptor.

2.3. **Edge-SPM visual histograms.** Lazebnik et al. [6] proposed Spatial Pyramid Matching (SPM), which divides each image into $2^l \times 2^l$ blocks in different scales $l = 0, 1, 2$, and computes the histograms of local features inside each block, followed by concatenating all histograms to represent the image. Let X and Y be two sets of vectors in d-dimensional. Then, the d-dimensional feature space is divided into hierarchical subsets at resolutions $l = 0, \ldots, L$, level l having 2^l cells along each dimension, so a total of D is equal to 2^{dl} cells.

Let H_x^l and H_y^l denote the histograms of X and Y at this resolution. And, the similarity of two sets of vectors is denoted by histogram intersection function as follows:

$$I\left(H_X^l, H_Y^l\right) = \sum_{i=1}^{D} \min\left(H_X^l(i), H_Y^l(i)\right) \tag{4}$$

The feature spaces are refined along with their increased scale, and the matching between features is more precise. The number of matches at level l also includes all the matches found at the finer level $l+1$. Therefore, the number of new matches located at level l is given by $I^l - I^{l+1}$, and the weight is set to $\frac{1}{2^{L-l}}$ with level l. The pyramid match kernel is denoted as:

$$K = I^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}} \left(I^l - I^{l+1}\right) \tag{5}$$

In this paper, the edge-SPM visual histograms are represented as an image by weighted multi-resolution histograms, which are obtained by iteratively splitting the training images into $2^l \times 2^l$ $(l = 0, 1, 2)$ blocks with variant scales, and computing the feature histograms by using edge-SIFT descriptors in each block, followed by generating edge feature histograms through splicing the sub-regions.

2.4. **SVM classifier.** SVM constructs a hyperplane to maximize the margin between different classes. The distance between the support vector and the classifier is indicated by the margins. SVM can transform a nonlinear separable problem into a linear separable problem with different kernel functions. In the experiments, the dimension of edge-SPM visual histograms is 6400. The edge-SIFT visual histograms are fed into SVM classifier. Finally, image classification results from SVM classifier.

3. **Experimental Results and Analysis.** Extensive experimental results are presented to evaluate the performance of the proposed method. A variety of experiments are conducted on three diverse datasets: Caltech101, Caltech256 and Comprehensive Cars (CompCars). The experiments are carried out on the Intel i5-3230M, 2.6GHz, RAM 8GB PC using MATLAB R2014b. Image classification is carried out with the support vector machine model – LibSVM [14] trained using Radial Basis Function (RBF) and Histogram Intersection Kernel (HI-K). In this experiment, accuracy rate acts as an essential evaluation index of classification performances. In general, the higher accuracy rate is, the better classification performances are. The accuracy rate (accuracy) is calculated as follows:

$$Accuracy = \frac{TP + TN}{P + N} \tag{6}$$

where TP (true positives) and TN (true negatives) are the number of correctly detected positive and negative samples respectively, $P + N$ is the total number of samples. Table 1

TABLE 1. Comparison of RBF and HI-K kernel function

Dataset	RBF	HI-K
Caltech-101	87.00% (87/100)	94.00% (94/100)
Caltech-256	78.89% (71/90)	87.78% (79/90)
CompCars-12	61.11% (110/180)	87.22% (157/180)
Mazda-8	60.83% (73/120)	85.00% (102/120)

lists the experimental results of the proposed algorithm using two libSVM kernels: RBF and HI-K on various datasets. As observed from Table 1, the performance of the HI-K outperforms the RBF, so this experiment selects HI-K as the kernel function for image classification.

3.1. **The Caltech-101 dataset.** The Caltech-101 [16] dataset contains 9144 images of 102 classes, including a background category. There exists significant deformation among different objects from the identical category. The proposed algorithm is tested on ten categories of objects selected from Caltech-101 dataset. For each category, the first 30 images are chosen as training images and the last 10 images as testing images. To verify the performance of the proposed algorithm, it is compared with BOW and SPM algorithm. The results of the classification are given in Figure 3. It can be derived from Figure 3 that the proposed algorithm achieves higher accuracy of the 6 classes of images than the BOW algorithm. For the two classes of ant and pigeon, the proposed algorithm is improved by 0.3 and 0.2 compared with the SPM algorithm. This is due to the fact that multi-scale Gabor filters are adopted, which optimize the optimum matching points. Moreover, average value of classification accuracy is 94% for our approach, which is 11% and 5% higher than the BOW and SPM algorithm respectively.

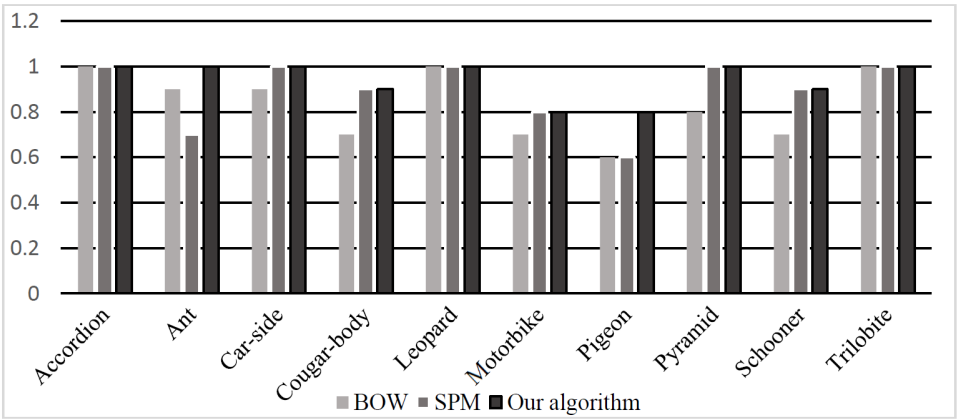


FIGURE 3. Comparison of three classification algorithms on Caltech-101 dataset

3.2. **The Caltech-256 dataset.** The Caltech-256 [17] dataset holds 30607 images falling into 257 classes, including a clutter category. It is an expansion of Caltech-101, and there are many higher intra-class variations and inter-class similarity. The sampled images contain nine categories. For each category, the first 30 images are chosen as training images and the last 10 images as testing images. Figure 4 reveals the classification performance of the proposed method, and it is compared with BOW and SPM algorithm on Caltech-256 dataset. The average values of classification accuracy for three methods are 87.78%, 85.60% and 86.67% respectively. The classification accuracy of each category is all higher than 70.00%, and the maximum value is 100.00%. These mean the collected template set has a certain discriminative ability, and the trained model has better robustness.

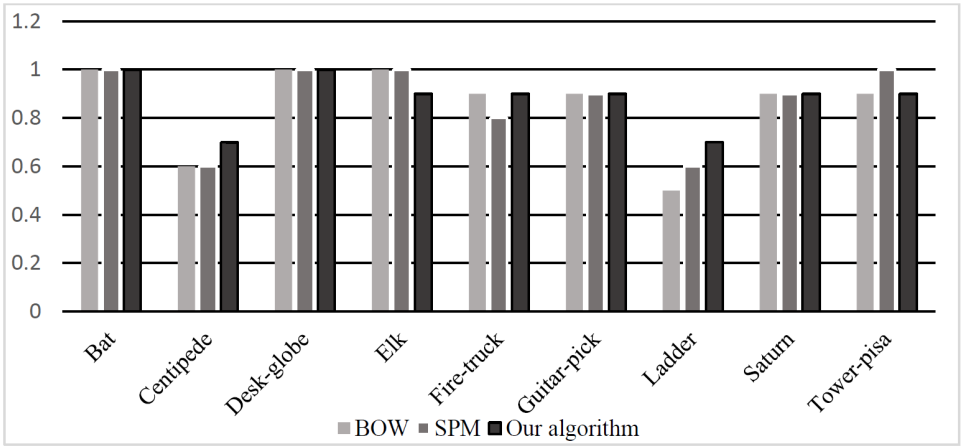


FIGURE 4. Comparison of three classification algorithms on Caltech-256 dataset

3.3. **The Comprehensive Cars dataset.** The CompCars [18] dataset contains 214,345 vehicle images from two scenarios: web-nature and surveillance-nature, which provides a comprehensive platform to validate the effectiveness of computer vision algorithms. The surveillance-nature data contains 50,000 car images. The web-nature dataset with 161 types of cars contains 1,687 car models. In a total of 136,727 images are the entire car and 27,618 images are the car parts images. First, 600 images with 12 car types are selected from web-nature dataset to evaluate the performance of the proposed algorithm, including Porsche, East-Wind, Fisher, Dacia, Zenvo and so on. Then, eight diverse models of the Mazda brand in the CompCars web-nature dataset are employed in this experiment. Sample images are listed in Figure 5. For each category, the first 35 images are chosen as training images and the last 15 images as testing images.

FIGURE 5. Eight various models of the Mazda brand in the CompCars dataset

As seen from Figure 6 the proposed algorithm achieves high accuracy on the discriminative categories, such as 93.00% accuracy for Fisker. The average classification accuracy is 87.22%. Table 2 illustrates the results of classification performance on eight diverse models of Mazda, in which the accuracy of only one vehicle model is less than 70.00% and their average accuracy is 85.00%. Since the contour similarity of the identical brand model is high, there are interferences to the classification. As a consequence, a new feature description method is required to explore for the classification of such types of images in the future work, in order to improve the classification discriminability.

4. **Conclusion.** In this paper, an improved image classification algorithm is presented by integrating multi-scale Gabor filters, edge-SIFT with SPM. The proposed method adopts

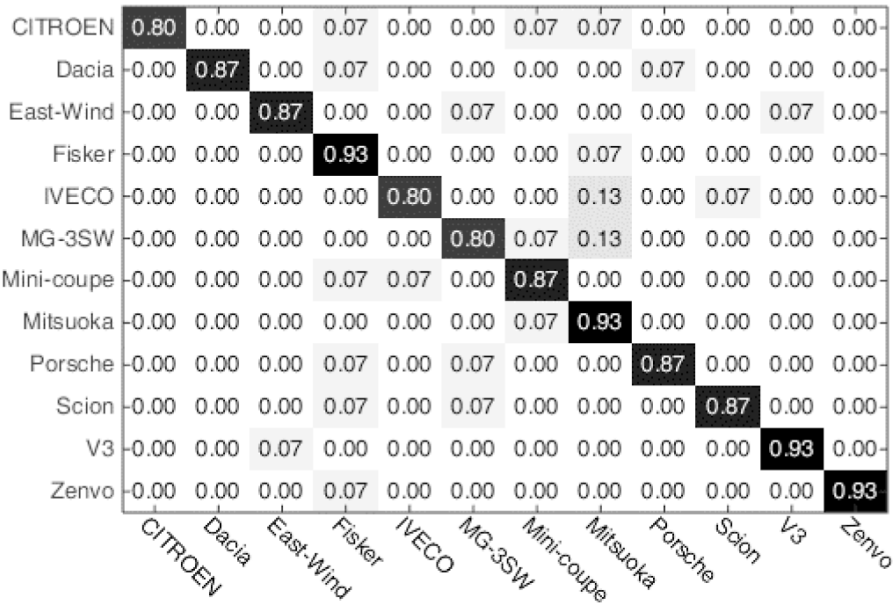


FIGURE 6. Confusion table for 12 car types from CompCars dataset

TABLE 2. Results by our method for eight diverse models of Mazda from CompCars dataset

Car types	Accuracy	Car types	Accuracy
Mazda-Hazumi	100.00%	Mazda-Axela	93.00%
Mazda3	93.00%	Mazda-zoom	87.00%
MazdaCX-7	87.00%	Mazda2	80.00%
MazdaCX-9	80.00%	Mazda8	60.00%

multi-scale Gabor filters to extract foreground information from edge image so as to reduce the redundancy information of the whole image. Moreover, the edge-SIFT instead of conventional dense SIFT algorithm is applied to SPM model. Experimental results on three distinct datasets indicate the proposed algorithm achieves better performance than BOW and SPM algorithm. Furthermore, the proposed method improves the accuracy and robustness of image classification significantly. In the future work, we focus on improvement of classification accuracy on low discriminability category.

Acknowledgement. This work was supported by the National Natural Science Foundation of China under Grant 61401127.

REFERENCES

[1] X. Peng, R. Yan and B. Zhao, Fast low rank representation based spatial pyramid matching for image classification, *Knowledge-Based Systems*, pp.14-22, 2015.

[2] Z. Y. Zhou, Y. C. Song and Z. F. Zhu, Scene categorization based on compact SPM and ensemble of extreme learning machines, *Optik*, pp.964-974, 2017.

[3] F. F. Li, R. Fergus and P. Perona, A bayesian hierarchical model for learning natural scene categories, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, pp.384-389, 2005.

[4] J. C. Yang, K. Yu and Y. H. Gong, Linear spatial pyramid matching using sparse coding for image classification, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Miami, FL, pp.1794-1801, 2009.

[5] L. X. Xie, Q. Tian and M. Wang, Spatial pooling of heterogeneous features for image classification, *IEEE Trans. Image Processing*, pp.1994-2008, 2014.

[6] S. Lazebnik, C. Schmid and J. Ponce, Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, pp.2169-2178, 2006.

[7] F. B. Silva, R. D. O. Werneck, S. Goldenstein, S. Tabbone and R. D. S. Torres, Graph-based bag-of-words for classification, *Pattern Recognition*, pp.266-285, 2017.

[8] D. G. Lowe, Distinctive image features from scale-invariant key points, *Int. J. Comput. Vis.*, vol.60, pp.91-110, 2004.

[9] L. F. Liu, Y. Ma and X. F. Zhang, High discriminative SIFT feature and feature pair selection to improve the bag of visual words model, *The Institution of Enigneering and Technology*, vol.11, pp.994-1001, 2017.

[10] M. Olgun and A. O. Onarcan, Wheat grain classification by using dense SIFT features with SVM classifier, *Computers and Electronics in Agriculture*, pp.185-190, 2016.

[11] P. Loncomilla and R. Solar, Improving SIFT-based object recognition for robot applications, *Image Analysis and Processing – ICIAP 2005*, Berlin, Heidelberg, pp.1084-1092, 2005.

[12] E. Fidalgo, E. Alegre and V. González-Castro, Compass radius estimation for improved image classification using edge-SIFT, *Neurocomputing*, pp.119-135, 2016.

[13] J. L. Zhou and Z. Zhang, Color image edge detection based on multi-scale Gabor filter, *Electronic Measurement Technology*, pp.49-52, 2016.

[14] C.-C. Chang and C.-J. Lin, LIBSVM: A library for support vector machines, *ACM Trans. Intelligent Systems and Technology*, pp.1-27, 2011.

[15] K. Li, C. Q. Zou and S. H. Bu, Multi-modal feature fusion for geographic image annotation, *Pattern Recognition Research*, vol.73, pp.1-14, 2018.

[16] F. F. Li, R. Fergus and P. Perona, Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories, *Comput. Vis. Image Understand.*, vol.106, no.1, pp.59-70, 2007.

[17] G. Griffin, A. Holub and P. Perona, Caltech-256 object category dataset, *California Inst. Technol.*, Pasadena, CA, USA, Tech. Rep., CNSTR-2007-001, 2007.

[18] L. J. Yang, P. Luo and C. C. Lou, A large-scale car dataset for fine-grained categorization and verifcation, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, Boston, pp.3973-3981, 2015.