

# A Novel Image Classification Algorithm Based on Word Bag Model and Feature Extraction

Zikang Tang<sup>1, a</sup>, Hao Zhang<sup>2, b</sup> and Fanlu Zhang<sup>3, c</sup>

<sup>1</sup>School of Engineering, Northeastern University, Boston MA , 02115, USA

<sup>2</sup>School of Geosciences and Info-Physics, Central South University, Changsha 410083, China

<sup>3</sup>Mechanical and Electrical Engineering Institute, Qingdao University of Science and Technology, Qingdao 266061, China

<sup>a</sup>bambo6taro@gmail.com, <sup>b</sup>sdzh422@outlook.com, <sup>c</sup>609361705@qq.com

**Keywords:** Image classification, feature extraction, Bag of Words, SIFT descriptor, SVM classifier.

**Abstract:** Image classification is based on the different characteristics of the image information to distinguish the images into different categories. It makes quantitative analysis of the image, interpreting the image or the image of each pixel region into a plurality of categories or in one, to replace the human visual interpretation. Traditional classification methods are usually based on single feature analysis. In this paper, a novel algorithm based on word bag model is described. This algorithm extracts a variety of image features at the same time, then establishes comprehensive access to multiple feature vectors, and finally gets more accurate image classification results. Bag of words model is a generalization of the traditional bag of words model. The core idea of the bag of words model method is that: all the words in the corpus statistics composed of words, words for each document statistics on the frequency of use, composed of the words frequency histogram to express this document. Corresponding to this model, we present the texture feature extraction and analysis steps.

## 1. Introduction

Digital images processing has been an integral part of people's life. There have been boosts as well as challenges since the recent years together with increasing technological advancements. One such challenging task is the efficiency to which a group of images or objects could be classified based on their traits, features, properties, etc. Image classification is an important part of image processing which helps to resolve this problem by using intelligent as well as feature based techniques.

Image classification problems are being applied in a wide range of applications including medical image classification, satellite image classification to classification of natural images. They are also being used for weather forecast and prediction, land terrain monitoring and registration applications [7], crop cultivation pattern identification, etc. The impact of image classification algorithms is tremendous in nature.

Feature extraction is the first and also a critical stage in image classification process as the number and type of feature extracted significantly impacts the classifier accuracy. Feature extraction is defined as the process of obtaining information from objects in the image to form feature vectors. These feature vectors are then used by classifiers to differentiate or segregate or identify or correlate the input unit with target output unit. It becomes easier for the classifier to classify between different classes by looking at these features. Prominently used feature extraction methods [12] are Template matching, Deformable templates, Unitary Image transforms, Graph description [4], Histograms [1], Contour profiles [21], Geometric moment invariants, Zernike Moments, Fourier descriptors, Gradient feature and Gabor features [2].

The second stage is classification which analyzes the feature vectors and classifies or puts them into suitable classes for ease of recognition. One challenge in classification is to predict the categories of the input image using its features. Various approaches for solving this problem such as k nearest neighbor (K-NN), Adaptive boost (Adaboosted), Artificial Neural Network (NN), Support Vector

Machine (SVM) have been discussed in this paper. For example, the k-NN classifier calculates the distance between the feature vector of the input image (unknown class image) and the feature vector of training image dataset and assigns the input image to the class among its k-NN. Also, the works in [17] presented LDA (Linear discriminant analysis) and PCA (Principal component analysis). Both of them extract features by projecting the original parameter vectors into a new feature space through a linear transformation matrix. But they optimize the transformation matrix with different intentions. PCA based techniques are more popular for classification problems, though they generally do not take into account the class information and are based solely on inputs. Another important classification algorithm is kernel based method, which is used to measure the membership of an image to a special image class. Due to the sheer number of images in each class, the training of kernel classifier may takes a long time. A new iterative algorithm to learn the kernel classifier has been discussed in [8], which is an efficient fast implementation of support vector machine (SVM). Experimental results indicate that the kernel based classifier is fast and robust, which makes the learning from a large number of images possible. Because the training images are input into the algorithm one by one, it does not takes much memory, and it is also an online algorithm.

One of such novel concepts that can potentially benefit the image classification algorithms is the Bag of words (BoW) model which can be basically applied as a feature generating tool. Bag of words model is a generalization of the traditional bag of words model. The core idea of the bag of words model method is that: all the words in the corpus statistics composed of words, words for each document statistics on the frequency of use, composed of the words frequency histogram to express this document. This paper presents a novel image classification algorithm which uses a BoW model to extract the feature and a state vector machine model to classify the data into its constituent classes. In this algorithm, BOW is predominantly used in text documents where the characters are extracted. Following the feature extraction, a classification tool is applied on the extracted characters to generate the classified results. A simple classification model is shown in Figure 1. The input cluster or object is listed into multi sets or bag of words followed by a classifier metric (e.g. distance), and classified into its constituent classes.

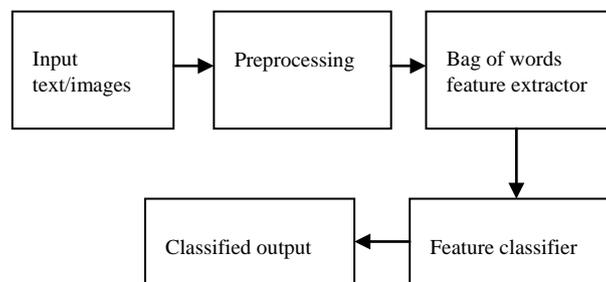


Figure 1. Schematic of image classification model

This paper is organized as follows. Section 2 describes the proposed methodology. Section 3 illustrates the experimentation results and analysis followed by conclusion in Section 4.

## 2. Proposed Work

The flow process of the proposed work is depicted in Figure 2 whose blocks are detailed in successive sections.

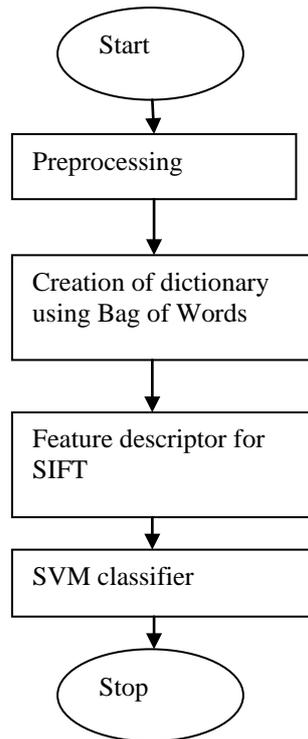


Figure 2. Flow process of proposed work

### 2.1 Bag of Words Feature Extraction

A Bag of Features (BoF) method is one that represents images as order-less collections of local features. The name comes from the Bag of Words representation used in textual information retrieval [6]. In case of textual document, bag of words is a normalized histogram of the total number of words in the document. The dictionary excludes non-informative words thus providing a reduction in data set size and hence the feature vector size. In the vector representation, each element is a term in the dictionary and the value of that element is the number of times the term appears in the document divided by the total number of dictionary words in the document. In the bag of words model, a visual vocabulary is constructed to represent the dictionary by clustering features extracted from a set of training images. The image features represent local areas of the image, just as words are local features of a document. Clustering is required so that a discrete vocabulary can be generated from millions (or billions) of local features sampled from the training data. Each feature cluster is a visual word. Given a novel image, features are detected and assigned to their nearest matching terms (cluster centers) from the visual vocabulary. The term vector is then simply the normalized histogram of the quantized features detected in the image. Features are extracted from training images and vector quantized to develop a visual codebook. The BoF term vector is a compact representation of an image which discards large scale spatial information and the relative locations, scales, and orientations of the features. Following feature extraction, a SIFT operator has been used for feature descriptor.

### 2.2 Support Vector Machine Classification

SVM has proven its efficiency over neural networks and RBF classifiers. The SVM approach seeks to find the optimal separating hyperplane between classes by focusing on the training cases that are placed at the edge of the class descriptors. These training cases are called support vectors. Training cases other than support vectors are discarded. This feature is very advantageous, especially for remote sensing datasets and more specifically for Object-based Image Analysis, where object samples tend to be less in number than in pixel based approaches. SVM uses an optimum linear separating hyperplane to separate two set of data in a feature space. This optimum hyperplane is

produced by maximizing minimum margin between the two sets. Therefore the resulting hyperplane will only be depended on border training patterns called support vectors.

The mapping function explained above is illustrated in Figure 3. The support vector machine operates on two mathematical operations:

- (1) Nonlinear mapping of an input vector into a high-dimensional feature space that is hidden from both the input and output.
- (2) Construction of an optimal hyperplane for separating the features discovered in step 1.

Variable definition:

1. Let  $x$  denote a vector drawn from the input space, assumed to be of dimension  $m_0$ .
2. Let  $\{\phi_j(x)\}$  for  $j=1$  to  $m_1$ , denote a set of nonlinear transformations from the input space to the feature space.
3.  $m_1$  is the dimension of the feature space.
4.  $\{w_j\}$  for  $j=1$  to  $m_1$  denotes a set of linear weights connecting the feature space to the output space.
5.  $\{\phi_j(x)\}$  represent the input supplied to the weight  $w_j$  via the feature space.
6.  $b$  is the bias
7.  $\alpha_i$  is the Lagrange coefficient
8.  $d_i$  corresponding target output

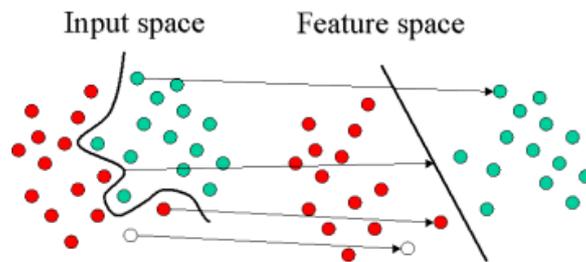


Figure 3. Illustration of SVM mapping function

### 3. Results and Discussion

To evaluate the effectiveness of the proposed LCMK method, we conduct extensive image classification experiments on 15-Scenes, Caltech 101/256. The scenes from the Caltech data base is depicted in Figure 4.

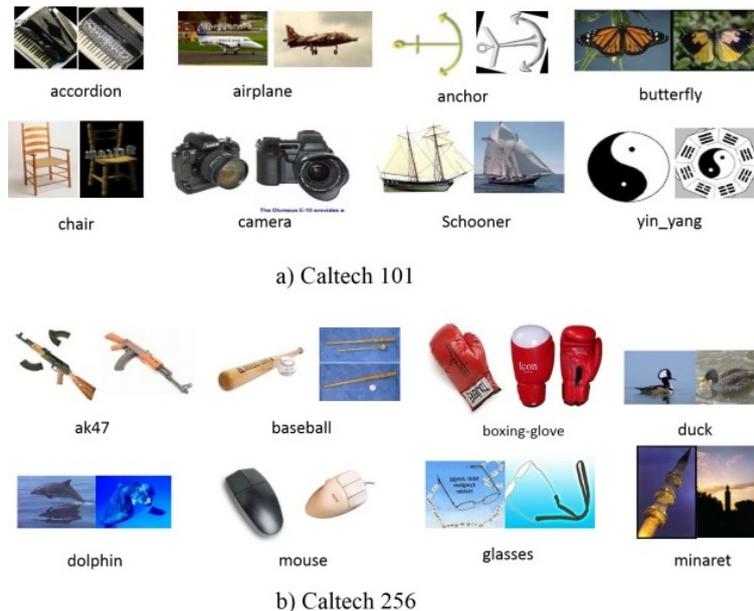


Figure 4. Sample test images from Caltech database

In the Caltech 101 dataset, there are 9144 images with 101 categories plus a ‘background’ category. The number of images in each category ranges from 31 to 800, and the image size is about 300×200 pixels. All 102 categories are used in the experiments. In the Caltech-256 dataset, there are 30,607 images from 256 categories plus a ‘background’ category. The number of images in each category ranges from 80 to 800. To obtain a good quality feature set, the input image is normalized and wiener filter used to remove the speckle noise. The Coefficients are used to reorder the column vector  $I_i$  of the images. In our method we use 2 level of decomposition C1, C2-1, C2-2, C2-3 coefficients to construct the feature matrix. Each coefficient is having 32x32 (total 1024) points. All the coefficients are arranged to make a column vector of 4028x1.

- The Feature image matrix  $I = [I_1, I_2, I_3 \dots I_P]$  is constructed from the coefficients column vector  $I_i$  where  $i$  represents the no of image.
- Feature matrix  $I$  is transformed to lower dimension subspace  $T_w$  using SVM.
- $T_w$  consists of Weight calculated for each image of the respective Dataset.

The pattern vectors (features) extracted from the images is given as input to the SVM classifier. Large database are required for the classifier to perform the classification correctly. In this system a sample of 180 bacterial images are collected and 120 images are used for training phase and remaining 60 images are used for testing. The classification accuracy and error rate is obtained by using the following formula:

$$Accuracy = \frac{\text{No.of correctly classified images}}{\text{Total number of images}} \quad (1)$$

and error is defined as

$$Error = \frac{\text{No.of incorrectly classified images}}{\text{Total number of images}} \quad (2)$$

Upon observing an unknown test image  $X$ , the weights are calculated for that particular image and stored in the vector  $W_X$ . Afterwards,  $W_X$  is compared with the weights of training set  $T_w$  using the Euclidean distance. If the distance does not exceed some threshold value, then the weight vector of the unknown image  $W_X$  is matched with the training dataset.

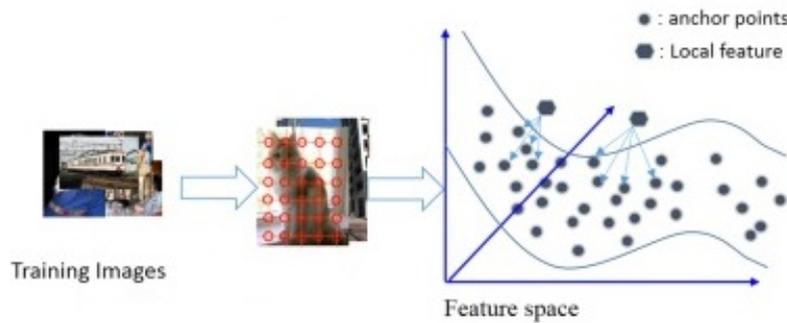


Figure 5 Mapping of input images onto the feature space

The optimal threshold value has to be determined empirically. All the algorithms are implemented in MATLAB 7.0.1, and executed on the Pentium-IV, 3.00GHz CPU with 2 GB RAM. The image data compression by means of PC extraction is accomplished by projecting each sample block of the data along the directions of the individual orthonormal eigenvectors of the covariance matrix of the image data. If the first few Eigen values of the covariance matrix contain most of the signal energy, the dimensionality of the data can be greatly reduced without losing much information by only retaining eigenvectors. In addition to its optimality in data reduction, which is not shared by other feature extraction schemes, it provides salient features of the data that are de-correlated. This property is ideally suited for classification purposes as these de-correlated features or components can be used

to train the classifier effectively. As shown in the table the feature matrix for 33 image becomes  $4096 \times 33 = 135168$  points. After applying SVM to the Feature Matrix I the weight matrix Tw becomes of size  $33 \times 33 = 1089$  Points. Thus for the classification of the image required to match only 1089 points ( $33 \times 33$ ) and not 135168 points. Thus the computational cost greatly reduced by applying SVM. The computation time is depicted in Figure 7. The performance is listed in Table 1.

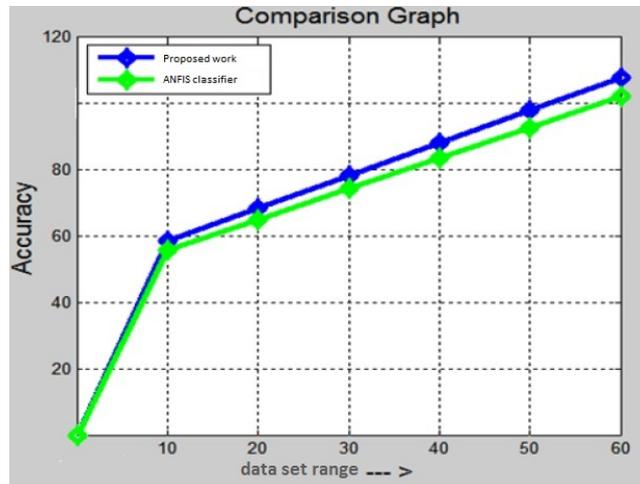


Figure 6 Classification accuracy of proposed work

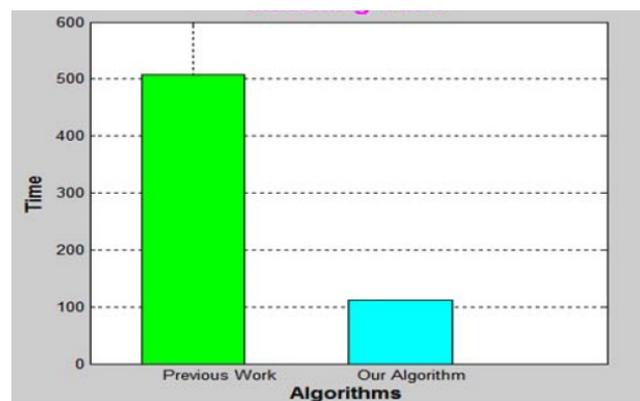


Figure 7 Computation time analysis

Table 1. Performance of proposed work (15scenes)

S. No.	Algorithm	Accuracy
1.	Proposed work	92.54%
2.	ANFIS classifier	90.69%
3.	SVM	86.12%

#### 4. Conclusion

In order to evaluate the accuracy improvements in the classification of high resolution imagery, the paper proposes novel algorithm which used a bag of words model for feature extraction and a SVM for classification. An evaluation conducted on image classification tasks using standard data

sets reveals the superiority of the proposed method which has been compared with ANFIS classifier with 3 layers and a conventional SVM classifier. There has also been a drastic reduction in the computation time and reduction in memory due to elimination of redundant data in the dictionary of BoW.

## 5. References

- [1] Pooja Kamavisdar, Sonam Saluja and Sonu Agarwal, "A survey on image classification approaches and techniques", *International journal of advanced research in computer and communication engineering*, Vol. 2, No.1, 2013.
- [2] Maneela Jain and Pushpendra Singh Tomar, "Review of image classification methods and techniques", *International journal of engineering research and technology*, Vol. 2, Issue. 8, 2013.
- [3] Zhiwu Lu, Liwei Wang and Ji Rong Wen, "Image classification by visual bag of words refinement and reduction", *Computer vision and pattern recognition*, 2015.
- [4] Li Li and Zhou Yan, "Image classification method based on improved bag of words", *Computer modelling and new technologies*, Vol. 18, pp. 242 – 246, 2014.
- [5] Chen Kai, Xiao Guo-qiang, and Pan Zhen, "Single-scale image classification employing Bag-of-Words model," *Application Research of Computers*, Vol. 28, No. 10, 2011.
- [6] Suralkhar, Karode and Pawade, "Texture image classification using support vector machine", *International journal of computer technology and applications*, Vol. 3, No. 1, pp. 71 – 75, 2012.
- [7] Shu-yi Zhang , Xiaorong Xue,Xi Zhang Kwang In Kim, Keechul Jung, Se Hyun Park, andHang Joon Kim "Support Vector Machines for Texture Classification, *IEEE Transactions IEEE Transactions on Pattern Analysis and MachineIntelligence*.Vol.24,No.11,November 2002.
- [8] Wang, Haoxiang, and Jingbin Wang. "An effective image representation method using kernel classification", *IEEE 26th International Conference on Tools with Artificial Intelligence, IEEE*, 2014.
- [9] D. Lu & Q. Weng, "A survey of image classification methods and techniques for improving classification performance", *International Journal of Remote Sensing*, Vol. 28, No. 5, 823-870, 2007.
- [10] Suresh Kashyap, Pooja Agarwal, Vikas Chandra Pandey and Suraj Prasad, "Soft computing based classification technique using KD99 data set for intrusion detection system", *International journal of advanced research in Electrical, Electronics and Instrumentation engineering*, Vol. 2, Issue 4, 2013.
- [11] Dubbey, Abhishek, "A study of classification techniques using soft computing", using MSER and improved spectral clustering" *EURASIP*, 2012.
- [12] Monireh Sheikh Hosseini and Maryam Zekri, "Review of medical image classification using adaptive neuro fuzzy inference system", *Journal of medical image sensing*, Vol. 2, No. 1, 2012.
- [13] M. Sharma and Mukherjee, "Fuzzy C means, ANFIS and genetic algorithm for segmenting Astrocytoma – A type of brain tumor", *International journal of advanced research in computer science and software engineering*, Vol. 3, No. 6, 2013.
- [14] Ada and Rajneet Kaur, " Feature Extraction and Principal Component Analysis for Lung Cancer Detection in CT scan Images" *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 3, No. 3, 2013.
- [15] Muhamad Naufal Mansor et al, "PCA- Based Feature Extraction and k-NN algorithm for Early Jaundice Detection", *International journal of soft computing and software engineering*, Vol. 1, No. 1, 2011.
- [16] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proceedings of Computer Vision and Pattern Recognition*, June 2003.
- [17] Xuechuan Wang and Kuldip Paliwal, "Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition", *Pattern recognition*, Vol. 36, pp. 2429 – 2439, 2003.
- [18] Le Hoang Thai, Tran Son Hai and Nguyen Thanh Thuy, "Image classificaiton using support vector machine and artificial neural network", *International journal of technology and computer science*, Vol. 5, pp. 32 – 38, 2012.

- [19] Bingquan Huo and Fengling Yin, "Research on Novel Image Classification Algorithm based on Multi-Feature Extraction and Modified SVM Classifier", *International journal of smart home*, Vol. 9, No. 9, pp. 103 – 112, 2015.
- [20] Yan Song, Ian Vince and Li Rong Dai, "Local coding based matching kernel method for image classification", *PLOS*, Vol. 9, No. 8, 2014.
- [21] Y. Huang, Z. Wu, L. Wang and T. Tan, "Feature coding in image classification, a comprehensive study", *Pattern Analysis and Machine Intelligence, IEEE Transactions on* vol. 36, no. 3, (2014), pp. 493-506