Image Enhancement using SVM

Shalini Goyal[1],Sudesh Nandal[2]
(Department of Electronics and Communication Engineering,
Bhagat Phool Singh Mahila Vishwavidyalaya, Khanpur Kalan, Sonepat, Haryana,India)

Abstract: In the field of agriculture, classification is one of the most popular assignments. In this paper, three different categories of rice grain images (basmati rice, brown rice and black rice) are classified by using Support Vector Machine. It is supervised machine learning classifier. There is an advantage of SVM is that it does not suffer any limitation of data dimensionality and limited samples. As per the study of comparison of rice image classification on various test sample ratios i.e. 0.5, 0.6, 0.7, 0.8, 0.9 and 1 gives the best result as average accuracy 86% is obtained from 0.9. This paper shows the comparative result of rice categories using different testing and training samples given to the classifier.

Keywords: Machine vision, support vector machine (SVM), bag of features, confusion matrix.

1. INTRODUCTION:

Rice is the widely consumed staple food in Asia. There are many variety of rice available in the world. In Asian region rice is mostly produced and consumed. After China, India has second rank in the production of rice. As demand is increasing for consuming rice, day by day then important to identify the variety of rice [1]. So, there is need of high quality and classified rice. The farmers are affected by this manual activity. Hence, these tasks require automation and develop imaging systems that can be helpful to identify rice grain images, rectify it & then being analyzed [2]. There are various methods available for classification of rice. To overcome this problem, machine vision and image processing techniques are effectively used for categorization of grain samples. A supervised learning classification technique is used by taking Support Vector Machine (SVM) as a classifier to classify the rice category which is required. So, this paper proposes a supervised learning method along with Bag Of Features and SVM for determining the variety of rice. The proposed method is formulated with the help of digital image processing tool in MATLAB. In this paper three different categories of rice images: basmati rice, black rice and brown rice are considered.

Data set/ Pre-processing:
The main step in any computer vision algorithm is data collection. There is no dataset available for rice image classification.
It was very difficult to collect data and pre-process it. The images for different classes of rice are obtained from different sources. For downloading images in bulk a tool Bulkimagedownloader is used.

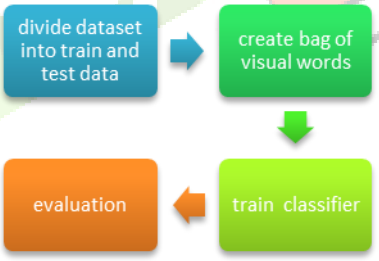

Fig.1 Work-Process

Colour images are used in our algorithm to extract features and codebook generation. All the gray-scale images are removed from the data. Images are not converted into binary because in this method color image is the main feature for classification. Initially images were of different sizes and format. All the images are resized into same size using MATLAB “imresize” function. The script used for pre-processing the data is “preprocessing.m”.
The dataset is divided into two categories:
1. Train set
2. Test Set

Bag of Features:
Bag of feature is the method that shows local features of images even they arranged order less [3]. Here is an example of an image with bag of features:

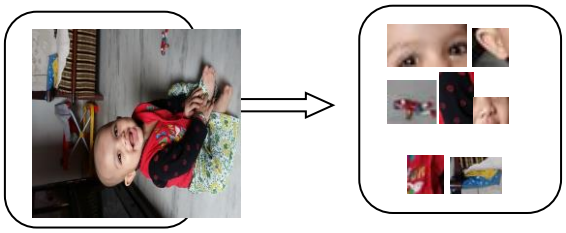

Fig.2 Bag of Visual Features

Feature Descriptors:
Local image features are described and also detected by using feature descriptor. Descriptors find features of images in such a way that it should not be affected by any perspective or any illumination. In some application it is not sufficient to extract only one type of feature to obtain the relevant information from the image data. Instead two or more different features are extracted, resulting in two or more feature descriptor at each image point. The set of all possible feature detector key points constitute a feature vector [4].

Clustering:
Clusters are created in order to reduce the number of descriptors in the codebook. What clustering actually done, it pick up the training images from the different categories and extract features or key points from each category for overall view of the different parts of the images or can say local features. The collection of local features in the form of feature vector which is a codebook or visual vocabulary. K-means algorithm aim to cluster n- features into k-clusters and return the k- cluster canters [5].

Classifier:
Classifier assigns bag of features representation of images to different classes [6]. This paper uses SVM as a trained classifier.

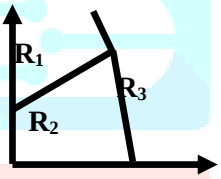

Fig.2 Data Classification

Support Vector Machine (SVM)
Vapnik and Chervonenkis introduced first Support Vector Machine (SVM) in the field of machine learning [7] in 1971 and since it has high degree of interest in the research field of machine learning research community [8].Machine learning is a process to teach the machine in natural way and makes an individual's to take decisions and predictions more accurately. In short, machine learning is "learn from experience"[9].

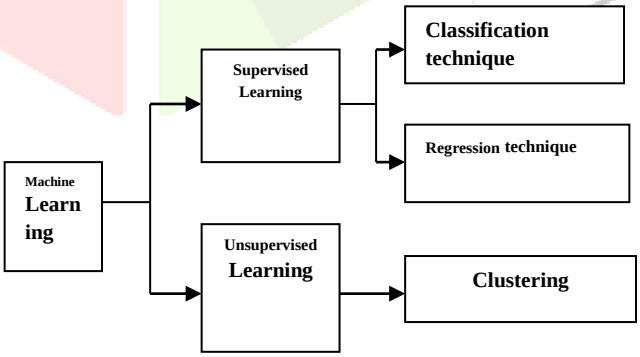

Fig. 4 Machine Learning Techniques [10]

There are two types of techniques in machine learning: Supervised and Unsupervised learning.

Supervised Learning: Supervised machine learning is based on known input and output data or in other way, evidence in the presence of uncertainty of the model and developed a predictive model. It also generates the reasonable predictions as per the response to new data after model has trained [9].To develops predictive models two techniques are used by supervised learning:

a. Classification technique
b. Regression technique

a. Classification technique: This technique predicts the discrete responses [3] i.e. classify input data into categories --- for example, whether a student is pass or fail.

b. **Regression technique:** This technique predicts the continuous responses i.e. classify input data into visual words (vocabulary) [3] --- for example, complete information about the student's result that in which subject it is pass or fail provided with marks.

Unsupervised Learning: Unsupervised learning is only based on input data and it grouped and interprets the data without labelled responses and also finds the hidden patterns or intrinsic structures in data.

Commonly clustering is used as unsupervised learning. Clustering is to find hidden patterns or grouping in data and also used for exploratory data analysis [10]. In this paper, k-means clustering is used to cluster all the data.

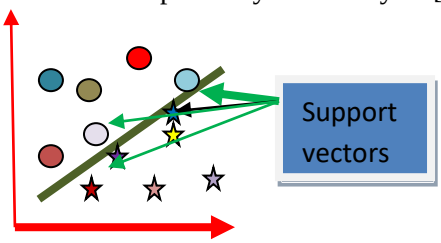

Fig.5 n-Dimensional Space showing Hyper-Plane [8].

SVM is based on machine learning algorithms. It commonly uses supervised machine learning algorithm that can be used for both classification and regression techniques. SVM applied mostly in classification problems.

Confusion matrix:

The classifier system gives the result as predicted classification and actual classification, both of this information enclosed in the matrix form that matrix is known as confusion matrix. The confusion matrix is generally evaluated to check classifiers performance [11].

2. IMPLEMENTATION:

This paper covers the details of the method used to build our classifier with specific references to the technique used, including an outline of the key functions in code in MATLAB using image processing tool.

Steps for proposed algorithm:

1. Initially image is converted from RGB color space to LAB color space. Unlike the RGB and CMYK colour models, LAB colour is designed to approximate the human vision. A bitmap image represented as LAB requires more data per pixel to obtain the same precision as an RGB or CMYK bitmap [5].
2. In this algorithm, two types of features are extracted from the image. Initially color features are extracted from the image. Color has an important role in the image classification. Black rice has very different color from brown rice or basmati rice.
3. HOG feature is also extracted from the image. These two features are combined to obtained better results. The extracted features are combined.
4. K- Mean clustering is used to obtain initial clusters from the extracted features. The algorithm groups the descriptors into mutually exclusive clusters. The resulting clusters are compact and separated by similar characteristics. Each cluster represents a feature or visual word.
5. The model is trained using a multiclass classifier using the error-correcting output codes (ECOC) framework with binary support vector machine (SVM) classifiers [12].
6. The classification model uses the bag of visual words to encode images in the image set into the histogram of visual words. The histogram of visual words is then used as the positive and negative samples to train the classifier [13].

Steps for implementation:

1. Download images form internet using bulkimagedownloader tool.
2. Arranged them in folder consist of three types of images as black rice, basmati rice and brown rice images.

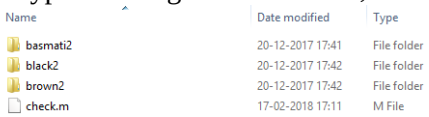

Fig.6 Rice Category File Folders

3. All these images having different size to make them of same size use a command resize, preprocessing is done.
4. Convert them from RGB color image to l*a*b* color images.
5. Using command counteachlabel no. of images are defined as,Basmati folder consists of 100 images, black folder consists of 97 images and brown folder having 90 images.
6. Now use the command spliteachlabel to split the images of each category with the ratio 3:7 defining the 30% training and 70% test images. Range of p from 0 to 1.

7. After that perform bag of features on training set images using function handler or custom extractor.

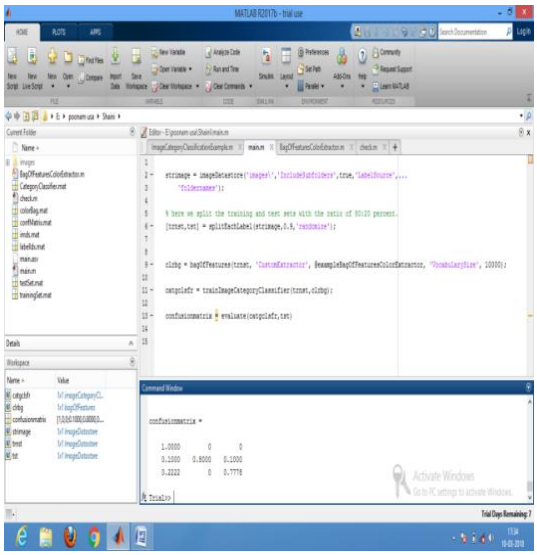

Fig.7 Command Used in the Program

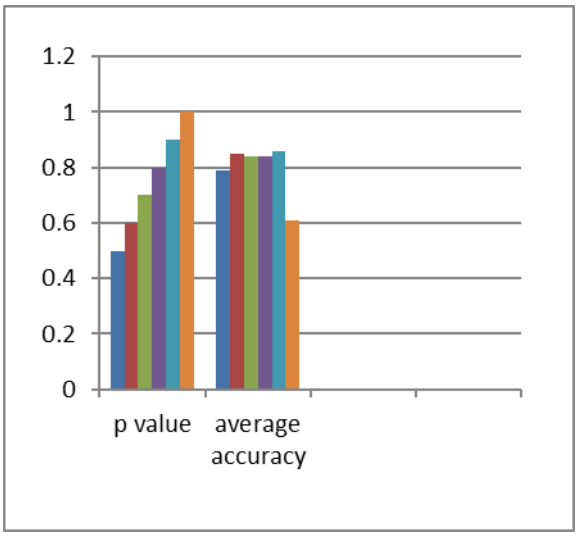

Fig.9 Comparison between p Values and Avg. Accuracy

8. K-means clustering is used here to create the vocabulary of 10000 visual words.
9. HOG color feature extractor is applied on the visual vocabulary.
10. Now SVM classifier is fully trained by training images.

Table1: Parameters with their values

Parameters	Value
Vocabulary Size	10000
Strongest Features	49767
Point Selection	Grid
Grid Step	[8, 8]
Block Width	[32, 64, 96, 128]

11. After that test images are given to SVM for classification.
12. The evaluation is done in the form of confusion matrix and average accuracy for all the different values of p.
13. Taking values p= 0.5, 0.6, 0.7, 0.8, 0.9 and 1.
14. Value p = 0.9 gives the best result with higher accuracy and classification among these values.

3. EXPERIMENTAL RESULTS:

The classification is performed over three categories of rice.

1. Basmati
2. Black
3. Brown.

Approx. 100 images are collected from various sources for each category. The training set consists of 90 images from each category.

The best result for p=0.9 value with highest accuracy 0.86 among all others. In this condition basmati images are 100% classified as basmati and also no other image is classified as basmati images.

Table2: Bag of Features values and average accuracy for different values of p.

S.N	Value of p	EFFI	NC	NSF	EF	Average accuracy (%)
1.	0.5	144	27648	10000	36864	79
2.	0.6	172	33177	10000	44032	85
3.	0.7	201	38706	10000	51456	84
4.	0.8	230	44238	10000	14746	84

5.	0.9	258	49767	10000	16589	86
6.	1	3	615	615	768	61

*EFFNI- Extracting features from no. of images **NSF- No. of strongest feature
***NC - No. of clusters ****EF- Extracted features

4. CONCLUSION:

This paper conclude the classification of rice images of three different varieties as black, brown and basmati rice images using the Bag Of Features and SVM (Support Vector Machine) as a trained classifier of training images then evaluate test images with the classification average accuracy of 0.79, 0.85, 0.84, 0.84, 0.86 and 0.61 for the **p** values of 0.5, 0.6, 0.7, 0.8, 0.9 and 1 respectively.

5. REFERENCES:

1. Prof. V.B.Raskar, Swapnil Thorat, Praviin mane, Eshwar Tak, "Rice Quality Analysis For Agmark Standard", International Engineering Research Journal (IERJ), Vol. 2 Issue 1, Page 290-292, 2016, ISSN 2395-1621.s
2. R.Kiruthika, S.Muruganand, Azha Periasamy, "Matching Of Different Rice Grains Using Digital Image Processing", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 7, July 2013.
3. Stephen O'hara And Bruce A. Draper, "Introduction To The Bag Of Features Paradigm For Image Classification And Retrieval", Arxiv:1101.3354v1 [Cs.Cv] 17 Jan 2011.
4. Benjamin Deavin, "Bag of Words: Automated Classification of Images", May 7, 2010.
5. K. Kavi Niranjana, M. Kalpana Devi, "RGB to Lab Transformation using Image Segmentation", International Journal of Advance Research in Computer Science and Management Studies, Volume 3, Issue 11, November 2015.
6. Cordelia Schmid, "Bag-Of-Features For Image Classification".
7. Durgesh K. Srivastava, Lekha Bhambhu, "Data Classification Using Support Vector Machine", Journal Of Theoretical And Applied Information Technology, 2005-2009.
8. Lei Shi, Qiguo Duan, Xinming Ma, Mei Weng, "The Research of Support Vector Machine in Agricultural Data Classification".
9. http://www.di.ens.fr/willow/events/cvml2011/materials/CVML2011_Cordelia_bof.pdf.Book of Mathworks, Introducing Machine Learning.
10. A. K. Santra, C. Josephine Christy, "Genetic Algorithm and Confusion Matrix for Document Clustering", IJCSI International Journal of Computer Science, Vol. 9, Issue 1, No 2, January 2012.
11. Sachine Shinde, Bharat Tidke, "Improved K-means Algorithm for Searching Research Papers", International Journal of Computer Science and Communication Network, Volume 4(6).
12. Devendra Dattatray Bhosale, "Use of Digital Image Processing for Grain Counting",International Journal of Advance Research in Computer Science vsdand Management Studies, Vol. 3, Issue 3, March 2015.
13. Iman Golpour, Jafar Amiri Parian and Reza Amiri Chayjan, "Identification and classification of Bulk Paddy, Brown, and White Rice Cultivars with Colour Features Extraction using Image Analysis and Neural Network", Czech J. Food Sci., Vol. 32, no.3, 280-287, 2014.
14. Nandini Sidnal, Uttam V. Patil, Pankaja Patil, "Grading And Quality Testing Of Food Grains Using Neural Network", (IJRET) International Journal of Research in Engineering and Technology, 2013.
15. Jaspreet Kaur, Vijay Kumar Banga, "Image Processing In Quality Assessment Of Pulses", IJCEM International Journal Of Computational Engineering And Management, Vol. 17 Issue 2, March 2014.