COMPARISON OF MODERN DESCRIPTION METHODS FOR THE RECOGNITION OF 32 PLANT SPECIES

Masoud Fathi Kazerouni, Jens Schlemper, and Klaus-Dieter Kuhnert

Institute of Real-Time Learning Systems, University of Siegen, Siegen, Germany

ABSTRACT

Plants are one kingdom of living things. They are essential to the balance of nature and people's lives. Plants are not just important to human environment, they form the basis for the sustainability and long-term health of environmental systems. Beside these important facts, they have many useful applications such as medical application and agricultural application. Also plants are the origin of coal and petroleum. In order to plant recognition, one part of it has unique characteristic for recognition process. This desired part is leaf. The present paper introduces bag of words (BoW) and support vector machine (SVM) procedure to recognize and identify plants through leaves. Visual contents of images are applied and three usual phases in computer vision are done: (i) feature detection, (ii) feature description, (iii) image description. Three different methods are used on Flavia dataset. The proposed approach is done by scale invariant feature transform (SIFT) method and two combined method, HARRIS-SIFT and features from accelerated segment test-SIFT (FAST-SIFT). The accuracy of SIFT method is higher than other methods which is 89.3519 %. Vision comparison is investigated for four different species. Some quantitative results are measured and compared.

KEYWORDS

SIFT, combined methods, HARRIS-SIFT, FAST-SIFT, feature extraction, feature detection, plant recognition

1. INTRODUCTION

Plants have a key role in the history as the actual resources for existence of life and stability on earth. They are key regulators of nature and lives as they provide most of the world's molecular oxygen and are the basis of most of the earth's ecologies, especially on land. Plant species provide vital ecosystem functions such as soil fertility and stability, water availability and pest control for sustainable agriculture, rangeland management and restoration. They have the unique ability to turn solar energy into chemical energy. Plants also play an important role for medical applications, alternative medicine and drugs. Due to their rich resources of ingredients, plants can be used in drug development and synthesis. Also plants provide people shelter, clothing, medicines, fuels, and the raw materials from which innumerable other products are made. Besides that these plants act a critical role in the development of human cultures around the whole world.

Study of plants is really necessary and indispensable because of their fundamental roles. Applications of plants in foodstuff, medicine, industry and agriculture lead to increase importance of plant recognition. There are different aspects of plants recognition. In context of agricultural applications, it is important distinguish and recognize different types of plants for various applications such as farm management. Also, plant recognition is very important for management of plants species whereas botanists can use it for medicinal purposes. Moreover, it can also be used for early diagnosis of certain plant diseases.

DOI : 10.5121/sipij.2015.6201

A plant recognition system can be used to identify plant species without need of any botany knowledge. Design of plant recognition system also contributes to fast classification, understanding and management of species. Common characteristics of plants could be useful in identifying them. Some characteristics of plants are stem, root, flower, fruit, and leaf. Due to some seasonal nature, inequality and changes of plants, the suitable characteristic is leaf. Some characteristics of leaves are form, edge, hair, and etc. Shape of leaf is the most important characteristic for plant species identification. These structures are parts of what make the leaves determinants, grow up and achieve certain patterns and shape, then stop. Other plant stems or root is parts, such as come determinant, usually will continue to grow, as long as they have the resources to do so. Although leaves vary in size, shape, and color, their typical characteristic is constant. The leaves are easy to find, collect and keep for digital images. Data collection process is approximately simple. Leaf of each plant carries useful information for classification of various plants, for example, aspect ratio, shape, color, and texture. Also one can easily transfer the leaf image to a computer and a computer can extract features automatically in image processing techniques.

Since plant species are determined by means of their leaves, main focus is on shape features and leaf contours. Curvature scale space (CSS) images were proposed by filtering leaf contours with Gaussian functions at different scales in [1]. Combining different shape features was also proposed for better classification performance [2], [3]. Shape context proposed for object recognition has also been used for leaf image classification [4], [5].

Methods such as SIFT [6] keypoints and descriptors have been used in object recognition and image classification widely [7]. In “[8]”, it was proposed to apply SIFT descriptors for flower image classification as SIFT descriptors are invariant to affine transformation.

There are a number of methods to measure the similarity of image content. BoW method is increasing widely used too [9], [10], and [11]. Impoverished representation of the data is offered by BoW model, because it ignores any spatial relations of features. Nevertheless, the model has proven to be very successful in the field of processing and analysis of natural language, mostly due to high discriminatory values of certain words in the text. It turns out that the same model can be successfully applied in the field of digital image processing.

The first important step towards the goal is selection of an available leaf image dataset. There are some datasets such as Flavia dataset [12], Leafsnap dataset, Intelengine dataset, and ImageCLEF dataset. The task was based on Flavia dataset.

Nowadays advanced pattern recognition, image processing, and machine learning methods and computer vision techniques are used for identification and recognition systems. With utilization of different advanced methodologies, we are able to implement an automatic plant recognition system with more functionality.

The present paper describes our automatic recognition system which is implemented by different high efficient methods and algorithms. The proposed approach consists of four phases: Image pre-processing, Feature detection and extraction, Training, and Testing.

Digital images of 32 different leaves are applied as dataset. The performance and accuracy of the tested methods are evaluated using the Flavia dataset.

The organization of the paper is as follows. Section II describes a general overview of architecture used by plant classifiers and related works, section III determines the proposed approach with discussions on extraction and computation of feature and classification schemes, section IV specifies details of the dataset and obtained experimental results, and section V moots the overall conclusion and scopes for next steps of research.

2. General Review

All over the world, there are plenty of plant species and subsequently a large volume of information of them. Development of a fast, reliable, and competent classification technique is necessary to handle the information.

Another important part of classification is implementing an effective method for representation of images. Visual contents of images are helpful to object recognition and do classification. Bag of words approach is used to measure similarities between images. A definition of the BoW model can be the "histogram representation based on independent features" [13]. BoW models are a popular technique for image classification inspired by models used in natural language processing [14]. The model ignores or downplays word arrangement (spatial information in the image) and classifies based on a histogram of the frequency of visual words. This method includes following three steps: feature detection, feature description, and visual words generation.

Detection of feature is a basic and important part for many applications such as image segmentation, image matching, object recognition, visual tracking in fields of image processing and computer vision. Feature detection is the process where automatically examine an image to extract features, that are unique to the objects in the image, in such a manner that an object could be detected based on its features in different images.

In computer vision and pattern recognition, feature extraction is one of main processing blocks. The primitive objectives of feature extraction are reducing the computational complexity of the subsequent process and facilitating a reliable and accurate recognition for unknown novel data. In other words, the goal of feature extraction is to yield a pattern representation that makes the classification trivial. As a challenge, feature extraction affects high efficient plant recognition. Extracting keypoints and descriptors are done by using algorithms such as SIFT and speeded up robust features (SURF) [15]. Different algorithms can be combined to investigate their effects. For instance, SIFT creates 128-dimensional vectors when SURF creates 64-dimensional vectors. By using the mentioned procedure, each image is a collection of vectors of the same dimension (128 for SIFT and 64 for SURF), where the order of different vectors is of no significance.

The last step is converting vectors represented features to visual words, which create a codebook (dictionary) [16]. The visual word "vocabulary" is established by clustering a large corpus of local features. One useful method is performing k-means clustering over the vectors. The centers of the learned clusters are visual words. Also the size would be the number of the clusters. Every feature is mapped to a specific visual word through the process of clustering and the image is represented by the histogram of the visual words.

The next process is training a classifier. SVM [17] was used to achieve the goal. Different SVMs will be used to find the best classifier. The final process is testing the classifier.

After training of a dataset, measurement of the algorithm accuracy is done by testing step. A set of testing images is used to find the accuracy of the used methods. Training dataset consists of 1255 leaf images and 32 classes.

Figure 1 shows general process for plant classification through leaf recognition.

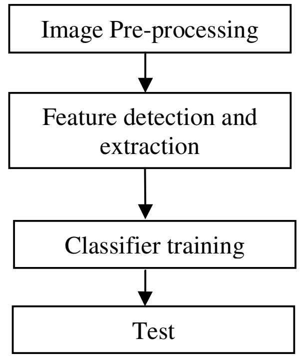


Figure 1. General architecture of plant classification system

3. LEAF RECOGNITION APPROACH

This approach is divided to four parts. These parts are image pre-processing, feature detection and extraction and classification. In the following, focus is on the details of method. The reason of using the method is classification accuracy and low computational cost.

3.1. Image Pre-processing

The leaf image is an RGB image. The first step is converting the RGB image to a gray scale image.

Eq. 1 is used for this purpose to convert an RGB pixel value to its gray scale value.

$$Gray\ scale = 0.299.R + 0.587.G + 0.114.B \qquad (1)$$

where *R*, *G*, and *B* correspond to the color of the pixel.

3.2. Feature Detection and Extraction

Feature detection is performed to do feature extraction for a set of labeled training images. An ideal detection should be possible even if some transformations, like scale and rotation, have been done.

Different feature detection methods use various schemes. Harris [18] method is rotation-invariant. Due to this characteristic, if the image is rotated, the same corners will be found. SIFT method is invariant to image scale, rotation, and highly distinctive. Therefore, another important characteristic is added to the first mentioned method as it is invariant to image scale. The last used method is Fast [19] method to do detection. It finds corners faster than other methods and is applicable for real time systems. The detected keypoints can be used for the next step.

After finding keypoints, descriptors are generated according to the areas surrounding interest points for a set of labelled training images. For this part, the used method is SIFT. A descriptor vector is computed for every keypoint. The dimension of the descriptor is 128. Although it seems to be high, lower descriptors than it do not perform the task as well as it. Also computational cost is another aspect of the process. Gained descriptors should be rich enough to be usable at the category level.

The next step is to create a vocabulary of features by clustering them. The vocabulary consists of cluster centres.

Descriptors of all images of the training set are taken and used to find the similarity between them. Similarity is determined by Euclidean distance between SIFT descriptors. Similar descriptors are clustered into K (here K equals 1000) number of groups. It is done by using Shogun's Kmeans class. The clusters are named visual words and they represent the vocabulary collectively. Each cluster has a cluster center which can be thought of as the representative cluster of all the descriptors belonging to that class. Cluster centres are found and used to group input samples around the clusters.

It is considerable that information of each detected keypoint in an image is mapped to a certain word through the clustering process and the image can be represented by the histogram of the words. The obtained vocabulary should be large enough to distinguish relevant changes in image parts, but not so large as to distinguish and recognize irrelevant variations such as noise.

3.3. Classifier Training

After vocabulary generation, each image is represented as a histogram of the frequency words that are in the image. This obtained vocabulary is about performing Object Categorization using SIFT descriptors of detected keypoints as features, and SVMs to predict the category of the object present in the image.

Now, BoW output is used to train a SVM.

A support vector machine constructs a hyper-plane or set of hyper-planes in a higher dimensional space, which can be used for classification, regression, or other problems. Intuitively, a good separation is achieved by the hyper-plane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger margin, the lower the generalization error of the classifier. The solution is optimal, which means that the margin between the separating hyper-plane and the nearest feature vectors from both classes (in case of 2-class classifier) is maximal. The feature vectors that are the closest to the hyper-plane are called support vectors, which means that the position of other vectors does not affect the hyper-plane (the decision function).

For given observations X, and corresponding labels Y which takes values +1 or -1, one finds a classification function

$$f(x) = sign(w^T x + b) \qquad (2)$$

where *w*, *b* represents the parameters of the hyper-plane.

Throughout training phase SVM needs a data matrix as input data and labels each one of samples as either belonging to a given class (positive) or not (negative), and then treats each sample in the matrix as a row in an input space or high dimensional feature space, where the number of attributes identifies the dimensionality of the space. SVM learning algorithm determines the best hyper-plane which separates each positive and negative training sample. The trained SVM can be deployed to perform predictions about a test samples (new) in the class.

SVM implementation is definitely important to get the best possible results for classification. Each type of SVM has some parameters. The used SVM type is υ-Support Vector Classification. It is N-class classification with possible imperfect separation. Parameter υ (in the range 0..1, the larger the value, the smoother the decision boundary) is used instead of C. Kernel type of SVM is Radial basis function (RBF). It is a good choice in most cases.

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \quad \gamma > 0 \qquad (3)$$

One of important parameters is cross-validation parameter, which is called kFold. The training set is divided into kFold subsets. One subset is used to test the model, the others form the train set. So, the SVM algorithm is executed kFold times. Here the parameter equals 10. An SVM with optimal parameters is trained. The method trains the SVM model automatically by choosing the optimal parameters C, gamma, p, nu, coef0, degree from parameters. Parameters are considered optimal when the cross-validation estimate of the test set error is minimal.

The final step of classification is testing. Prediction of different existed leaves in testing dataset is performed and the results of the combined methods are acquired.

4. EXPERIMENTS

Flavia dataset is comprised of 32 different plant species. SVM classifier with RBF kernel is used on the dataset. The images are used to train classifier of each method. For test data, 648 images are used to test the accuracy of the proposed method. The used machine is Intel® Core™ i7-4790K, CPU @ 4.00 GHz, and installed memory (RAM) 16.0 GB. Investigation of the methods is illustrated in below.

Three combined methods are used to obtain their results on the dataset. These methods are SIFT, HARRIS-SIFT, and FAST-SIFT.

Table 1 shows the accuracy result of three performed methods.

Table 1. Accuracy of classification.

Used method	**Accuracy of classification (percent)**
SIFT	89.3519
FAST-SIFT	81.9444
HARRIS-SIFT	80.4012

As it is shown, the best result is obtained by using SIFT. The reason lies in the characteristics of this method. In SIFT algorithm, scale, rotation, and contrast invariant features are extracted. Also SIFT keypoints are extrema in a difference-of-Gaussian (DoG) scale pyramid. Once potential keypoints locations are found, they have to be refined to get more accurate results. They used Taylor series expansion of scale space to get more accurate location of extrema, and if the intensity at this extrema is less than a threshold value (0.03 as per the paper), it is rejected.

DoG has higher response for edges, so edges need to be removed too. For this purpose, a concept similar to Harris corner detector is used. They used a 2×2 Hessian matrix (H) to compute the principal curvature. For edges in Harris corner detector, one eigen value is larger than the other. An orientation is assigned to each keypoint to achieve invariance to image rotation. A neighbourhood is taken around the keypoint location depending on the scale, and the gradient magnitude and direction is calculated in that region. An orientation histogram with 36 bins covering 360 degrees is created. The highest peak in the histogram is taken and any peak above 80% of it is also considered to calculate the orientation. It creates keypoints with same location and scale, but different directions. It contributes to stability for achieving a good result.

Now keypoint descriptor is created. A 16×16 neighbourhood around the keypoint is taken. It is divided into 16 sub-blocks of 4×4 size. For each sub-block, 8 bin orientation histogram is created. So, a total of 128 bin values are available. It is represented as a vector to form keypoint descriptor. In addition to this, several measures are taken to achieve robustness against rotation, and etc.

Harris detector is used to detect corners. The detected points of this method did not have the level of invariance required to obtain reliable image matching. The method has been widely used for some specific computer vision applications, but the accuracy of it is less than other methods.

Fast detector performs faster than other methods and improvement in the computational speed is considerable, but it is not very robust to the presence of noise. Since high speed is achieved by analyzing the fewest pixels possible, the detector's ability to average out noise is reduced.

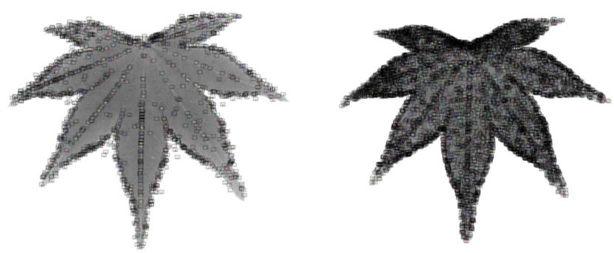
Number of keypoints is calculated for four species of the dataset. The difference of four species is complexity due to human vision. They are labeled as: Simple, approximately simple, approximately complicated, complicated. For SIFT method, the complicated one has the maximum number of keypoints while the minimum number of keypoints belongs to the simple one.

The number of keypoints for SIFT and FAST-SIFT are calculated (Table 2).

Table 2. Number of keypoints.

Number of keypoints in method	**Simple leaf**	**Approximately simple leaf**	**Approximately complicated leaf**	**Complicated leaf**
SIFT	103	154	663	1639
FAST-SIFT	528	322	1165	6180

Keypoints are calculated for one leaf by SIFT and FAST-SIFT methods. The images are shown in Figure 2.



(a) (b)

Figure 2. (a) Representation of keypoints for SIFT method. (b) Representation of keypoints for FAST-SIFT method

The FAST method finds thousands of keypoints, while other used methods find only hundreds. In above, it is mentioned that detection with FAST method generates some noise keypoints. Large number of keypoints, keypoints mixed up with noisy keypoints, cause decrement of accuracy. Detected keypoints of SIFT are enough and also accurate to have a good result.

For performing the FAST-SIFT method, the required time is less than other methods. Needed test time is calculated for the methods and evaluated per image. Table 3 shows the obtained values.

Table 3. Test time per image.

Method	Needed test time per image (ms)
SIFT	780.43
FAST-SIFT	610.39
HARRIS-SIFT	771.87

Also, it is discovered that performance of FAST-SIFT method is better than HARRIS-SIFT method. One possible reason is large number of detected keypoints. However, increase of number of keypoints leads to detect noisy keypoints, the reason of better performance in comparison of HARRIS-SIFT lies in increase of keypoints. So it is anticipated to have a better result for FAST-SIFT.

In general, each detected pixel has a very little information. Descriptors use relationship of pixels and model them to have better information. The detected keypoints affect the results of plant recognition. SIFT method has the best result between the used methods.

Effects of varying some parameters on final error for three methods using RBF kernel are investigated. Investigation is done on Nu and Gamma parameters.

The parameter nu is an upper bound on the fraction of margin errors and a lower bound of the fraction of support vectors relative to the total number of training examples. The value of nu parameter is between 0 and 1. To do the experiment, gamma parameter is kept fixed and equals 1.0. As it is shown in Figure 3, error of the methods is increased when this parameter increases. By use of SIFT method, increase of nu parameter has less influence on results. SIFT has a better robustness against varying nu parameter.

The effect of changing the gamma parameter on error is shown in Figure 4, while nu parameter is held fixed at 1.0. There is a direct relationship between increase of gamma parameter and error's increase. In comparison with nu parameter, the impact of gamma increase on final results is less than nu parameter.

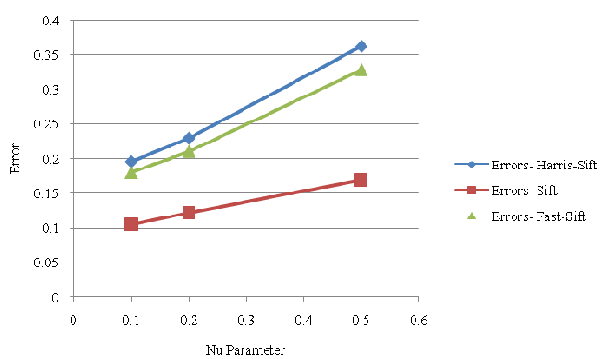


Figure 3. Variation of Nu parameter for used methods

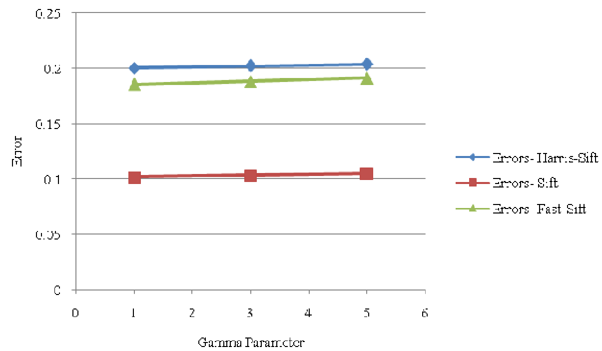


Figure 4. Variation of Gamma parameter for used methods

An interesting matrix is constructed, which is named confusion matrix. The confusion matrix is one $n \times n$ matrix ($n = 32$ in our case) containing information about the actual classification results (in its columns) and different category labels through the classification (in its rows).

After calculation of confusion matrix, precision and recall values are calculated for each label of the used methods.

$$Precision_i = \frac{M_{ii}}{\sum_j M_{ji}} \qquad (4)$$

$$Recall_i = \frac{M_{ii}}{\sum_j M_{ij}} \qquad (5)$$

Generally, precision is the fraction of events where we correctly declared i out of all instances where the algorithm declared i. Conversely, recall is the fraction of events where we correctly declared i out of all of the cases where the true of state of the world is i.

Precision is a measure of result relevancy, while recall is a measure of how many truly relevant results are returned. In both Figure 5 and Figure 6, the minimum values of the methods belong to HARRIS-SIFT method. It is predictable as it has the least accuracy percentage between three methods. Values variation of SIFT method is less than other methods in both figures. In comparison, FAST-SIFT method variation is in middle of the other methods and has the second rank in these figures. After precision and recall measurements, the sequence is SIFT, FAST-SIFT, and HARRIS-SIFT.

Surrounded area is another concept to compare the methods. A high area under the curve represents both high recall and high precision, where high precision relates to a low false positive rate, and high recall relates to a low false negative rate. In Figure 5 and Figure 6, SIFT method has larger areas and proves better performance of the method. Both high scores show that the method is returning accurate results (high precision), as well as returning a majority of all positive results (high recall).

The relationship between recall and precision can be observed for each method.

For SIFT method, the minimum value of recall is less than 0.6, while the minimum value of precision is more than 0.6. Variation of values in precision is less than the variation in recall. In Figure 7, these variations are shown.

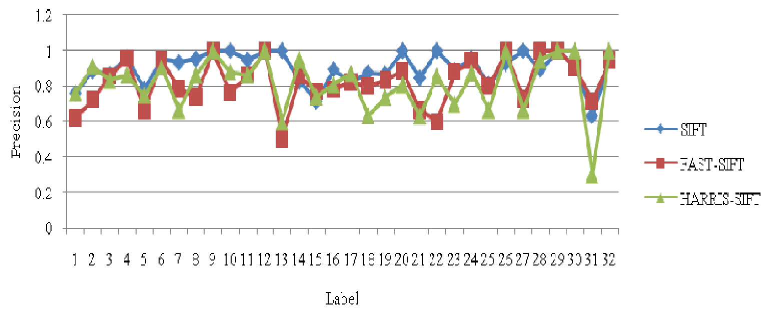


Figure 5. Precision measurement for SIFT, FAST-SIFT, and HARRIS-SIFT methods

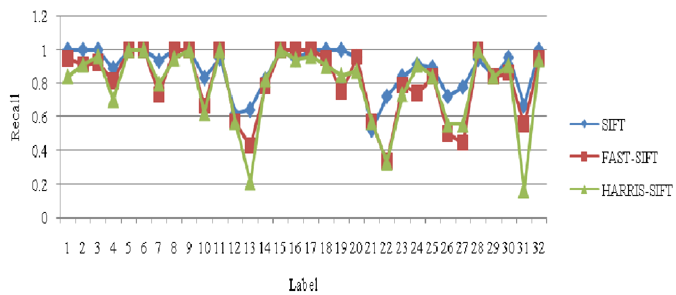


Figure 6. Recall measurement for SIFT, FAST-SIFT, and HARRIS-SIFT methods

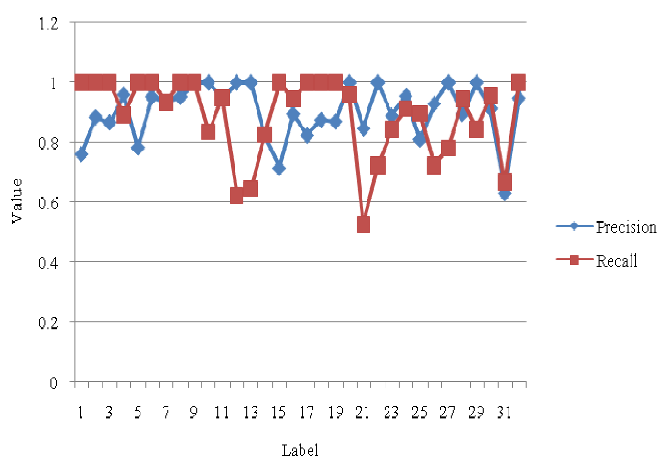


Figure 7. Precision and Recall values for SIFT method

Also, more labels have values near the maximum and most of values are in the highest interval, which is [0.8, 1].

In Figure 8, precision and recall values are defined for FAST-SIFT method. Intervals of variation are larger than SIFT method. The minimum value of precision is 0.5. After investigation of minimum value of recall, it is found that the value is less than 0.4 and equals 0.3333.

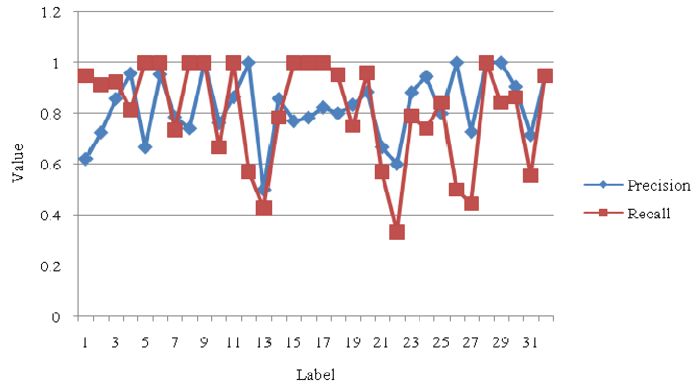


Figure 8. Precision and Recall values for FAST-SIFT method

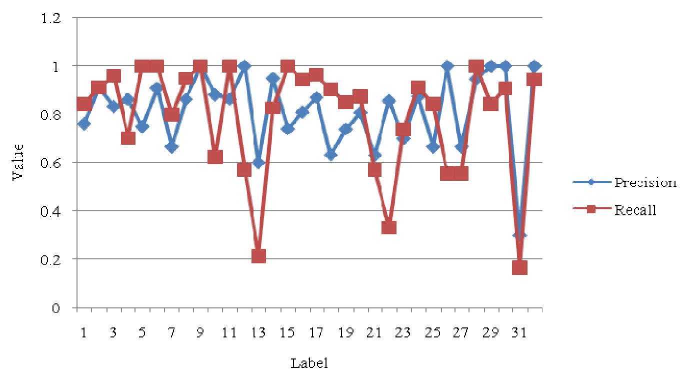


Figure 9. Precision and Recall values for HARRIS-SIFT method

The Figure 9 shows precision and recall of HARRIS-SIFT method. The minimum values of precision and recall are 0.3 and 0.166667. The values intervals of precision and recall are larger than other two methods. It is expectable, because the accuracy of this method is lower than other methods. Decrease of its values is the evidence.

5. CONCLUSIONS

In this paper, SIFT method and two combined methods are taken into consideration for plant recognition and classification. Accuracy measurement and efficiency of each method are described. The methods were tested on Flavia dataset. Experimental results are also compared with some quantitative results and discussed according to human vision for four different species. Experiments on the dataset, demonstrate that SIFT method has the best performance between proposed methods.

The proposed work can be applied for other algorithm, SURF. Some other combinations of different methods of detection and extraction of features can be used for next steps.

APPENDIX

Obtained confusion matrix is shown for each method. Table 4 is confusion matrix of SIFT method, while Table 5 and Table 6 belong to FAST-SIFT and HARRIS-SIFT methods.

Table 4. Confusion matrix of SIFT.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	19	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
1	0	23	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
2	0	0	26	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
3	2	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	18	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
5	0	0	0	0	0	20	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
6	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
7	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	1	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	13	0	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	2	0	0	0	0	0	0	0	1	0	9	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0
19	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0
20	2	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	6	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	1	0	0	0	0	0	0	0
22	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	16	0	0	0	0	0	0	0	0	0
23	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0
24	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	13	0	1	0	0	1	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	14	0	0	0	2	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0
28	0	0	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	16	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	21	0	0
30	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	1	0	0	12	0
31	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	18

Table 5. Confusion matrix of FAST-SIFT.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	18	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2	0	0	24	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
3	0	1	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	18	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
5	0	0	0	0	0	20	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
6	0	3	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	2	0	2	0	0	0	0	16	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	1	0	0	0	0	0	0	0	0	12	0	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	2	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	1	2	0	2	0	0	0	0	0	0	0	0	1	0
13	0	0	0	0	0	0	0	0	0	0	1	0	0	18	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	1
18	0	0	0	0	1	0	0	0	0	0	1	0	3	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0
19	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0
20	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	6	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	1	0	0	0	0	0	2	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	3	0	15	0	0	0	0	0	0	0	0	0
23	0	2	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0
24	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	2	0	0	0	0	0
25	4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	3	0	0	0	0	9	0	0	0	0	0	0
26	0	0	0	0	1	0	0	0	0	2	1	0	1	0	0	0	0	0	0	1	0	0	1	0	2	0	8	0	0	0	1	0
27	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	18	0	0	0	0
28	0	0	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	16	0	0	0
29	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	19	1	0
30	0	0	1	0	0	0	0	5	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	10	0
31	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	18

Table 6. Confusion matrix of HARRIS-SIFT.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0
1	0	21	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
2	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
3	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	18	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
5	0	0	0	0	0	20	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
6	0	1	0	0	0	0	12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	2	1	0	0	0	0	0	15	0	0	0	0	0	0	0	1	0	0	2	0	2	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	1	0	0	0	0	0	0	0	0	12	0	1	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1	0	1	0	0	1	0	0	0	0	1	0	3	0	0	0	0	0	1	3	0	1	0	0	0	0	0	0	0	0	2	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	19	1	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	17	0	0	0	0	0	0	0	1	0	0	0	1	0
19	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	6	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	1	0	0	0	0	0	0	1	0
22	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	1	0	14	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	21	0	0	0	0	0	0	0	0
24	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	2	0	0	0	0	0
25	3	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	1	0	1	0	0	0	0	10	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	3	0	3	0	10	0	0	0	0	0
27	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	18	0	0	0	0
28	0	0	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	16	0	0	0
29	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	2	0	0	20	0	0
30	0	0	1	0	0	0	0	2	0	1	0	0	0	0	0	1	0	4	0	0	1	0	0	0	4	0	0	1	0	0	3	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	17

ACKNOWLEDGEMENTS

The authors would like to thank you the authors of Flavia dataset for making it available publicly.

REFERENCES

[1] S. Abbasi, F. Mokhtarian, and J. Kittler, "Reliable Classification of Chrysanthemum Leaves through Curvature Scale Space," in *Proceedings of the First International Conference on Scale-Space Theory in Computer Vision*, Utrecht, The Netherlands, Jul. 1997

[2] Z. Wang, Z. Chi, and D. Feng, "Shape Based Leaf Image Retrieval," *IEE Proceedings of Vision, Image, and Signal Processing*, vol. 150,pp. 34–43, 2003

[3] X.-Y. Xiao, R. Hu, S.-W. Zhang, and X.-F. Wang, "HOG Based Approach for Leaf Classification," in *Proceedings of the International Conference on Intelligent Computing (ICIC)*, Changsha, China, pp. 149–155, Aug. 2010

[4] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, Apr. 2002

[5] H. Ling and D. W. Jacobs, "Shape Classification Using the Inner-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 286–299, Feb. 2007

[6] Lowe, D. G., "Object Recognition from Local Scale-Invariant Features," in *Proc. Of the International Conference on Computer Vision*, pp. 1150—1157, 1999

[7] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005

[8] M.-E. Nilsback and A. Zisserman, "Automated Flower Classification Over a Large Number of Classes," in *Indian Conference on Computer Vision, Graphics and Image Processing*, Dec. 2008

[9] G. Csurka, C. Dance, L. Fan, J. Williamowski, and C. Bray "Visual Categorization with Bags of Keypoints," in *Proc. Of SLCV Workshop, European Conference on Computer Vision (ECCV'04)*, pp. 59-74, Prague, 2004

[10] F. Perronnin, C. Dance, G. Csurka and M. Bressan, "Adapted Vocabularies for Generic Visual Categorization," in *Proc. Of ECCV'06*, Graz, 2006

[11] J. Winn, A. Criminisi, and T. Minka, "Object Categorization by Learned Universal Visual Dictionary," *IEEE International Conference on Computer Vision*, pp.1800-1807, Beijing, 2005

[12] S. Wu, F. Bao, E. Xu, Y.-X. Wang, Y.-F. Chang, and Q.-L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," in *Signal Processing and Information Technology, 2007 IEEE International Symposium on*, pp. 11 -16, Dec. 2007

[13] L. Fei-Fei, R. Fergus, and A. Torralba. "Recognizing and Learning Object Categories, CVPR 2007 short course", 2007

[14] Sivic, Josef, "Efficient Visual Search of Videos Cast as Text Retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4. IEEE, pp. 591-605, Apr. 2009

[15] Bay, H., A. Ess, T. Tuytelaars, and L. Van Gool. "SURF:Speeded Up Robust Features," *Computer Vision and Image Understanding (CVIU)*,vol. 110, no. 3, pp. 346–359, 2008

[16] T. Leung and J. Malik. "Representing and Recognizing the Visual Appearance of Materials Using Three-Dimensional Textons," *International Journal of Computer Vision*, vol. 43, no. 1, pp. 29–44, 2001

[17] Christopher J. C. Burges. "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, vol. 2, no. 2, pp. 1-43, 1998

[18] C. Harris, M. Stephens, "A Combined Corner and Edge Detector," in *Proceedings of the Fourth Alvey Vision Conference*, pp. 147-151, 1988

[19] E. Rosten, T. Drummond. "Machine Learning for High-Speed Corner Detection," in *European Conference on Computer Vision*, vol. 1, pp. 430–443, May 2006