

Efficient Modern Description Methods by Using SURF Algorithm for Recognition of Plant Species

¹Masoud Fathi Kazerouni, ²Jens Schlemper and ³Klaus-Dieter Kuhnert

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

¹masoud.fathi@uni-siegen.de; ²schlemper@fb12.uni-siegen.de; ³kuhnert@fb12.uni-siegen.de;

ABSTRACT

Plants are one of one of most valuable natural resources. There would be no life on earth without plants. They unlike humans and animals manufacture their own food by photosynthesis. In every food chain, the plants occupy the first position and lead the chain as source of food. Environment and climate are largely interlinked with plants. Rainfall, humidity and temperature are influenced by presence of plants. Cutting down plants also imbalance the environment which will indirectly affect human life. Even the economic importance of plants is also quite large to mankind. Plants are great contributors of economy. Many countries rely on agriculture as one of the main source of revenue. Other benefits of plants are significant applications in different fields. Medical and agricultural applications are just some instances of plants application.

Plant recognition can be done by using unique characteristic parts of plants. The used part is leaf. Shapes of leaves are useful to do plant recognition and find the species. Bag of words (BoW) and support vector machine (SVM) methods are applied to recognize and identify plants species. Visual contents of images are used and four steps are performed: (i) image preprocessing, (ii) BoW, (iii) train, (v) test. Three combined methods are used on Flavia dataset. The proposed approach is done by Speed-up robust features (SURF) method and two combined method, HARRIS-SURF and features from accelerated segment test-SURF (FAST-SURF). The accuracy of SURF method is higher than other applied methods. It is 92.28395 %. In addition to visual comparison, some quantitative results are measured and compared.

Keywords: Text Component; SURF; combined methods; HARRIS-SURF; FAST-SURF; feature extraction; feature detection; plant recognition.

1 Introduction

Over 3,000 million years ago, the first living-organism which resembled a plant appeared. It was a blue-green algae which lived in the sea and can still be found in the water today. When the plants made their first appearance on Planet Earth, the atmosphere was not appropriate for all oxygen breathing creatures. The air was made out of carbon dioxide, a gas which to us is deadly. Then photosynthetic plants came along and slowly over several million years, cleaned the atmosphere and filled it with oxygen. Ever since early man rubbed two sticks together to make fire, plants have played a vital role in the history of mankind. Over time, utilization and application of plants are increased in different fields. For a few decades at the beginning of the 20th century, growing and learning from a garden of medicinal plants was part of the pharmacy curriculum [1]. A sharply increasing interest in the commercial growth of medicinal plants led to more attention to plants.

Medicinal application of plants are still increasing and playing a fundamental role in this field. Researchers are trying to promote profitable and sustainable agrifood, fibre and horticultural industries, developing new plant products and improving natural resource management. Also recent developments of agriculture have changed traditional agriculture and added new applications and techniques to increase agricultural production and protect environment. Application and existence of plant are undeniable in human life and environment. Hence, recognition and distinguish of plant species are increased.

Determination of plant species is a challenge for non-botanists. Recognition process may take a long time and be grueling. For instance, traditional farmers do not have enough knowledge of different plant species that may grow between their products. Existence of unwanted plants like weed, which is a plant in the wrong place, can damage products of a farm. Due to plant industry and increase of new plant products, necessity and need of a plant recognition system are increased. In context of biological applications, health of plants is very important during a research and accurate and fast recognition of plant disease are significant. Accurate identification of a cultivated plant can be very helpful in knowing how it grows (e.g., size shape, texture, etc.) as well as how to care and protect it from pests and diseases. Also, a plant recognition system would let non-professionals to obtain scientific knowledge of botanists. These are only some aspects of plant recognition system's need.

Since last decade, plant recognition has become a popular topic in computer vision and image processing. Design and implantation of a plant recognition system are urgent needs. An efficient study is vital to achieve this goal. The characteristics of plants give us clues as to how some plants are similar to and different from others. Classification of plant species depends upon common and unique features that are used to identify plants by their characteristics. Scientists and plant experts have collected data on numerous plant species from studying the plants in their natural habitats and recording information about their characteristics in scientific literature and databases for future reference. Although plants can be identified by the shape and color of their fruits, flowers, stems, color of stems, and the seeds they contain, but leaves are the most useful characteristics for recognition and identification of plants. Plants leaves contain important amount of information. Leaf features include the color, texture, shape, size and orientation of the plant's leaves on the stem, end etc. Some features like texture and color may vary with different environmental conditions and time. So they are not useful because of their variations. Shape of leaves is the best feature of plants to do classification and identification of various species. By using digital images of leaves, it would be easy to collect and use them for further process. Information of leaves will be used for performing the methods.

Plant recognition was done by using leaf shapes-a case study of the Acer family in [2]. A polygon approximation was used to classify different species of Acer family. In [3], simple shape features as centroid-contour distance (CCD) curve, eccentricity and angle code histogram (ACH) were applied to recognize plants of 140 Chinese medical plants. Douglas-Peucker approximation algorithm was used to extract leaf shape features in [4]. After that, basic geometric features and digital morphological features were used in [5]. In [6], snakes technique with cellular neural networks (CNN) was employed to do recognition. Also, shape context has been used for doing leaf classification [7], [8]. A widely used keypoint based detector and descriptor is SURF [9]. Hessian-matrix approximation of integral image is used by SURF. In "[10]", a comparative study of SURF and Scale-invariant feature transform (SIFT) [11] has been performed.

Bag of features [12] representation is very popular for content based image classification due to good performance and simplicity. SIFT and SURF can be used in image categorization. Generally,

better performance and efficiency of training and classification depend on better representation and clustering of features.

Bag of Words (BoW) model [13, 14, 15, 16, 17] is originated from natural language processing tasks and information retrieval. For image analysis, a visual analogue of a word is used in the BoW model, which is based on the vector quantization process by clustering low-level visual features of local regions or points, such as color, texture, and so forth [18]. Representation of it uses image patches as visual words. One important benefit of using BoW model is increase of accuracy in classification as a result of building a large dictionary.

In [19], a plant recognition algorithm is proposed which is implemented with three different modern description methods. Now, the purpose is to have a better feature extraction that leads to higher accuracy of the algorithm.

The first requirement of designing an automatic plant classification is to find and select a suitable dataset. The used dataset is Flavia dataset [20]. It is comprised of 32 different plant species. The dataset is divided into two datasets, training dataset and test dataset. The training dataset consists of leaf images of training dataset for in [19]. Also the test dataset is built with the same images of test dataset in [19]. All proposed methods are tested by the dataset.

The used methods are originated from computer vision and machine learning fields. To achieve the goal, advanced and modern methodologies are investigated to design and implement the desired plant recognition system with high efficiency and accuracy.

Using useful and advanced methods and techniques of pattern recognition, image processing, and machine learning lead to implementation of an automatic plant recognition system with acceptable accuracy. Identification and recognition of different plant species can be used in different fields of science and industry.

The present paper explains an automatic recognition system which is implemented by modern description methods and algorithms. The approach includes four phases: Image pre-processing, Feature detection and extraction, Train, and Test.

This paper is organized as follows: Section II deals with general overview of the used architecture for plant classification and related works, Section III introduces the applied algorithm with explanation on detection and extraction of feature and the schemes of classification, Section IV describes the results and the experiments conducted using the proposed methods, and Section V presents the conclusion and next steps of research.

2 General Review

Shapes of plant leaves can help botanists and biologist to identify and recognize various plants species as they have an important amount of information. Because of large number of plants species all around the world, an automatic plant recognition system is helpful to classify them. Advancement and progress of computer vision can help to develop an accurate and reliable system for this purpose.

An important part of a recognition system is to handle information. Leaves of plants have significant information. It is essential to extract information for recognition systems. Developing an efficient and fast system needs to find an effective method for extracting information of leaves helps.

In image classification, an image is classified according to its visual content. The principle basis is contents of the image. Classification is based on the similarity of the image contents. The image contents are described via image features. Visual contents are needed for object recognition and identification. Detecting keypoints with rich information should be done as a basis part. It can be done automatically by using different detection methods and represented by descriptors [21]. Then, keypoints are grouped into a number of clusters with those with similar descriptors assigned into the same cluster. They would be shown by one single visual word. So, all keypoints will be mapped to a limited number of visual words.

A popular and widely used method for feature representation is Bag-of-Words (BoW). It is used for document representation in information retrieval. This methodology was first proposed in the text retrieval domain problem for text document analysis [22], and it was further adapted for computer vision concepts and applications [22]. For image analysis, a visual analogue of a word is used in the model, which is based on the vector quantization process by clustering low-level visual features of local regions or points, such as color, texture, and so forth.

Detection of local interest points and features has an important role in different image processing concepts and fields. It is also the first step of BoW model. Automatic detection of features is performed to detect meaningful points in images. They are quite unique to the objects of images. Due to this process, any object will be detected based on its own features in each image. Several well-known region detectors have been described in the literature [23, 24].

The next step of the model is to compute feature descriptors for the detected features. It is a main part of the procedure. In general, feature extraction involves reducing the amount of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. In this specific purpose, the importance of it is not deniable. Extracting keypoints and descriptors is done by using SURF algorithm. In [19], SIFT algorithm was used for combined methods. Now SURF algorithm is used and its influence on results is investigated. SURF creates 64-dimensional vectors or 128-dimensional vectors when SIFT creates only 128-dimensional vectors. The order of difference vectors is of no significance. The implemented SURF is used by 64-dimensional vectors.

After detection and extraction of features, the next step is based on vector quantization. For this task, K-means clustering is performed and the number of visual words generated is based on the number of clusters. Clustering algorithm is used to find the centers of clusters of the feature descriptors and visual vocabulary is the collection of the visual words. For each feature descriptor in an image the nearest visual word from the vocabulary is assigned to it. The distribution of visual words in the image is represented as histogram.

After the BoW feature is extracted from images, it is entered into a classifier for training or testing. SVM [27] is used to train the classifier. Using a training dataset is necessary in this step. After generating a classifier, the test process is done. In this step, a test dataset is needed too. Experiments are done with different combined methods and the results are obtained. The number of classes is 32.

Figure 1 shows a general scheme of plant recognition system.

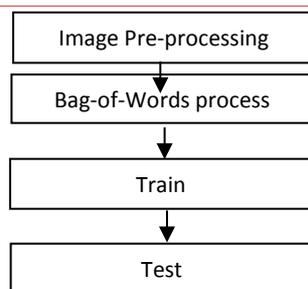


Figure 1: General scheme of plant recognition system.

3 Leaf Recognition Approach

3.1 Image Pre-processing

The leaves images of dataset should be converted to gray scale images to be used as input images. Conversion RGB images to gray scale images is the first step to do preprocessing and the algorithm.

The following equation, Equation (1), is the used formula to convert RGB pixel values to gray scale values.

$$\text{Gray scale} = 0.299.R + 0.587.G + 0.114.B \quad (1)$$

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

3.2 Bag of Words

Local feature extraction involves interest point detection and computation of descriptors in region surrounding those interest points.

The current step is to perform feature detection. Features can be detected manually or, preferably, it can be detected automatically using some specific techniques. Feature must be prominent, easily detectable and spread over the whole image. Feature detection method should have good localization accuracy and should not be sensitive to the assumed image degradation. The used method should be able to detect features regardless of image deformation. Transformation such as scale, and rotation should be detectable.

Each feature detection method has its own characteristics. For instance, Harris [25] method is a rotation-invariant method for feature detection. When rotation is occurred, finding the same corners is its important characteristic. It is one of the used methods. SURF method is a speed-up version of SIFT method and is rotation and scale invariant. In SIFT, Lowe approximated Laplacian of Gaussian (LoG) with Difference of Gaussian (DoG) for finding scale-space. SURF goes a little further and approximates LoG with Box Filter. One big advantage of this approximation is that, convolution with box filter can be easily calculated with the help of integral images. And it can be done in parallel for different scales. Also SURF method relies on determinant of Hessian matrix for both scale and location. As a robust local feature detection method, it works much faster than SIFT method. For feature detection, another used method is FAST (Features from Accelerated Segment Test) [26] method which was proposed by Edward Rosten and Tom Drummond in their paper "Machine learning for high-speed corner detection" in 2006 (later revised it in 2010). FAST method is a weapon choice to do detection faster. Its utilization is undeniable in real time systems. This method finds a lot of features and it is one of important characteristics of it. Generally, the goal of a descriptor is to provide a unique and robust description of an image feature, e.g. by describing the intensity

distribution of the pixels within the neighborhood of the point of interest. Most descriptors are computed thus in a local manner; hence, a description is obtained for every point of interest identified previously. The SURF descriptor is based on the similar properties of SIFT, with a complexity stripped down even further. The first step consists of fixing a reproducible orientation based on information from a circular region around the interest point. The second step is constructing a square region aligned to the selected orientation, and extracting the SURF descriptor from it. A descriptor vector is computed for every keypoint. The dimension of the descriptor is 64 and it is less than SIFT method with the dimension of 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. SURF method has lower dimension, higher speed of computation and provides better distinctiveness of features. Obtained descriptors will be used to find similarity between different images.

Sets of keypoint descriptors are used to represent images. Then, vector quantization (VQ) technique is used to cluster the keypoint descriptors in their feature space into a large number of clusters. K-means clustering is applied and each keypoint is encoded. K-means algorithm is used to assign of points to their closest cluster centers and computation of the cluster centers. Each cluster is a visual word and represents a special pattern by the keypoints in the cluster. Now a visual word vocabulary is generated. It contains local patterns in images. After mapping the keypoints to visual words, each image can be represented as a bag of visual words. Thus, a novel image's features can be translated into words by determining which visual word they are nearest to in the feature space (based on the Euclidean distance between the cluster centers and the input descriptor). As a point, vocabulary size should be large enough to recognize changes in image parts. Increase of size may lead to distinguish irrelevant variations and quantization artifacts.

3.3 Classifier Train

After generation of vocabulary, each image is represented by a histogram of how often the local features are assigned to each visual word. The representation is known as bag-of-visual-words in analogy with the bag-of-words (BOW) text representation where the frequency, but not the position, of words is used to represent text documents. Applying a classifier is the next step of the approach. The classifier is typically a support vector machine (SVM) [27] which is it often known to produce state-of-the-art results in high dimensional problems. The output of BoW method is used to do classification and train a SVM.

In general, SVMs are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. SVM performs classification tasks by constructing hyper-planes (support vectors) in a multidimensional space that separates cases of different class labels. It supports both estimating the relationships between variables (regression) and data classification and can handle multiple continuous and categorical variables. Also, SVM benefits from two good ideas: maximizing the margin and the kernel trick. These good ideas can guarantee high testing accuracy of classifiers and overcome the problem about curse of dimensionality. In other words, SVM constructs hyper-planes either in input space or in feature space from a set of labeled training dataset. The hyper-plane will try to split the positive samples from the negative samples. The linear separator is commonly constructed with maximum distance from the hyper-plane to the closest negative and positive samples. Intuitively, this causes correct classification for training data which is near, but not equal to the testing data.

Consider a binary classification problem with N training samples (data). Each sample is indicated by a tuple (x_i, y_i) and $(i = 1, 2, \dots, N)$, where $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ corresponds to the attribute set for the i th sample. Conventionally let $y_i \in \{-1, 1\}$ and it is considered as its class label. The decision boundary of a linear classifier can be written as follows:

$$w^T x + b = 0 \quad (2)$$

Where w is weight vector and b is a bias term.

The SVM implementation goal is to define a decision boundary that is maximally far away from any data point. For training process, SVM needs an input matrix and labels each 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. Determination of the best hyper-plane to separate each positive and negative training sample is one of important steps. The obtained and trained SVM will be used to perform predictions of the test dataset and find the labels.

SVM is an efficient algorithm for classification due to high accuracy, nice theoretical guarantees regarding overfitting, and with an appropriate kernel. There are several types of SVM. The type of used SVM is u -Support Vector N -class classification. Kernel type of SVM is Radial basis function (RBF), because support vector machines employing the kernel trick do not scale well to large numbers of training samples or large numbers of features in the input space, several approximations to the RBF kernel (and similar kernels) have been devised. Like other types of SVM, it has some parameters. Instead of C parameter, u parameter is used, which is in the range $0..1$. When it increases, the decision boundary becomes smoother. The training set is divided into k fold subsets. k Folds is cross-validation parameter and the SVM algorithm is executed k fold times. This parameter equals 10 for the approach. Automatic training is done by choosing optimal SVM parameters such as γ , ρ , ν , coef0 , degree from parameters. When the cross-validation estimate of the test set error is minimal, parameters are optimal.

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

32 different plant species of Flavia dataset are applied to do experiments. Each method is tested with the following machine and the accuracy of them is evaluated. The machine is Intel® Core™ i7-4790K, CPU @ 4.00 GHz, and installed memory (RAM) 16.0 GB. Investigation of the experiments is done on the dataset by SVM classifier with RBF kernel. Also, test dataset comprised of 648 images. The procedure of experiments is illustrated in below.

SURF, HARRIS-SURF, and FAST-SURF are proposed as three methods to obtain results on the dataset and evaluate accuracy of tests.

Table 1 shows the accuracy result of three performed methods. SURF method has the highest accuracy between the used methods and the methods that are used in [19]. FAST-SURF method has a higher accuracy when it is compared to SIFT method in [19].

Table 1: Accuracy of classification.

Method	Accuracy of Classification
SURF	92.28395
FAST-SURF	89.6605
HARRIS-SURF	87.1914

As it is shown, the maximum accuracy belongs to SURF method. SURF is a speed-up version of SIFT. It is invariant with regard to scale, orientation, and illumination. In SIFT, Lowe approximated Laplacian of Gaussian (LoG) with Difference of Gaussian (DoG) for finding scale-space. SURF goes a little further and approximates LoG with Box Filter. One important advantage of this approximation is that, convolution with box filter can be easily calculated with the help of integral images. And it can be done in parallel for different scales. Also the SURF relies on determinant of Hessian matrix for both scale and location.

To assign orientation, SURF uses wavelet responses in horizontal and vertical direction for a neighborhood of size $6s$. Adequate Gaussian weights are also applied to it. Then, they are plotted in a space. The dominant orientation is estimated by calculating the sum of all responses within a sliding orientation window of angle 60 degrees. Interesting part is that, wavelet response can be found out using integral images very easily at any scale.

For feature description, SURF uses Wavelet responses in horizontal and vertical direction (again, use of integral images makes things easier). A neighborhood of size $20s \times 20s$ is taken around the keypoint where s is the size. It is divided into 4×4 sub-regions. For each sub-region, horizontal and vertical wavelet responses are taken and a vector is formed. The represented vector gives SURF feature descriptor with total 64 dimensions. Lower the dimension, higher the speed of computation and matching, but provide better distinctiveness of features.

For more distinctiveness, SURF feature descriptor has an extended 128 dimension version. The sums of dx and $|dx|$ are computed separately for $dy < 0$ and $dy > 0$. Similarly, the sums of dy and $|dy|$ are split up according to the sign of dx , thereby doubling the number of features. It doesn't add much computation complexity.

Using sign of Laplacian (trace of Hessian Matrix) for underlying interest point is one of important improvement. It adds no computation cost since it is already computed during detection. The sign of the Laplacian distinguishes bright blobs on dark backgrounds from the reverse situation. In the matching stage, features comparison is only possible if they have the same type of contrast. This minimal information allows for faster matching, without reducing the descriptor's performance. Totally, SURF adds a lot of features to improve the speed in every step.

Corner detection can be done by Harris detector. Harris points are preferable when looking for exact corners or when precise localization is required. It basically finds difference in intensity for a displacement of (u, v) in all directions. This method is not reliable, because detected points of the method do not have the required level of invariance for image matching. Although the performance of the method is less than other methods, it has been used widely for different computer vision applications.

To have a faster detection, FAST detector is applied. It is based on the accelerated segment test (AST), which is a modification of the SUSAN [28] corner detector. FAST is not robust to the presence of noise and high level noise. Beside this disadvantage, it is many times faster than other existing corner detectors. Also, it has high levels of repeatability under large aspect changes and for different kinds of features.

The second experiment is calculation of number of keypoints for four different species. These species are selected from the dataset. Complexity of the species is specified by human vision. The labels are named Simple, approximately simple, approximately complicated, complicated. The number of keypoints is calculated for SURF and FAST-SURF methods (Table 2).

Table 2: Number of keypoints.

Number of keypoints in method	Simple leaf	Approximately simple leaf	Approximately complicated leaf	Complicated leaf
SURF	161	47	350	829
FAST-SURF	528	322	1165	6180

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



Figure 2: (a) Representation of key points for SURF method. (b) Representation of key points for FAST-SURF 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, key points mixed up with noisy keypoints, cause decrement of accuracy. Detected keypoints of SURF are enough and also accurate to have a good result.

The needed time for performing any method is computed and shown in Table 3. The needed time per image is calculated. FAST-SURF method has the minimum required time and it is the fastest one. In comparison with SIFT method [19], SURF method needs less time for the same dataset. The needed time for HARRIS-SURF method is lower than HARRIS-SIFT method in [19].

Table 3: Test time per image.

Method	Needed test time per image (ms)
SURF	445.268
FAST-SURF	345.512
HARRIS-SURF	528.756

By using SURF method, faster computation is obtained without sacrificing performance. In short, SURF adds a lot of features to improve the speed in every step. Analysis shows it is faster than SIFT while performance is comparable to SIFT. SURF is good at handling images with blurring and rotation, but not good at handling viewpoint change and illumination change.

Each detected point has a little information. Descriptors use their relationship to have enough information and a useful model. Fast-SURF method finds more keypoints than HARRIS-SURF method. Thus, more information is obtained which helps to have a better performance. Although FAST algorithm detects some noisy keypoints, increase of the number of keypoints leads to perform better. FAST-SURF method has repeatability and this characteristic has a good influence on its performance. Due to the characteristics of HARRIS and FAST algorithm, it is anticipated to obtain better results when FAST-SURF method is used.

The detected keypoints affect the results of plant recognition. In SURF method, number of keypoints is not too high or low. It is completely enough to extract information and have an accurate method with acceptable results.

SVM parameters can be changed to investigate their effects on the methods. Two SVM parameters are chosen to consider their variation effects on final error of the methods. The selected parameters are Nu and Gamma parameters.

Nu parameter and C parameter are equivalent regarding their classification power, but Nu parameter has a more meaningful interpretation. This is because Nu presents an upper bound on the fraction of training samples which are errors (badly predicted) and a lower bound on the fraction of samples which are support vectors. Nu parameter has a value between 0 and 1. Nu parameter is changed, while Gamma parameter is held fixed and equals 1.0. When Nu parameter increases, error of methods is increased. It is shown in Figure 3. Increment of Nu parameter has the least influence on SURF method in comparison to other methods. Robustness of SURF method against Nu parameter variations could be a helpful characteristic.

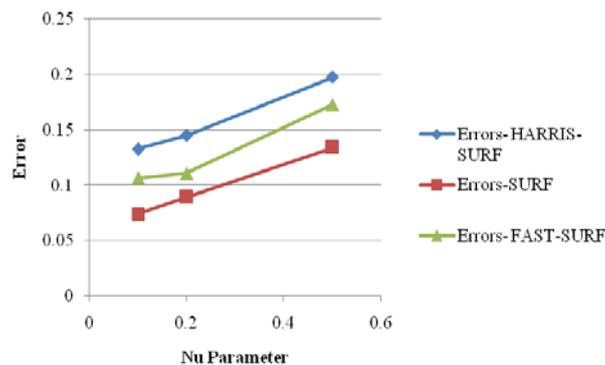


Figure 3: Variation of Nu parameter for used methods.

The next investigated parameter is Gamma. Gamma parameter variations are performed to consider the effects on final error of each method. *Gamma* parameter defines how far the influence of a single training example reaches, with low values meaning far and high values meaning close. For this experiment, Nu parameter is kept fixed at 0.1. Increase of Gamma parameter causes error's increment. It is shown in Figure 4. The diagrams are ascending, but the values of slopes are not high. The minimum slope value belongs to SURF method that indicates the robustness of the method in this experiment. In general, increase of Gamma parameter has less impact than increase of Nu parameter.

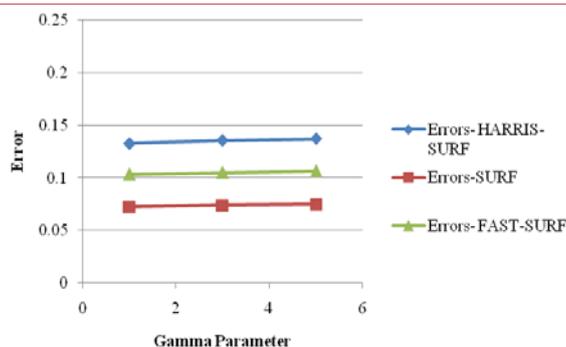


Figure 4: Variation of Gamma parameter for used methods.

Visualization of the performance of methods is performed by construction of confusion matrix. A confusion matrix contains information about actual and predicted classifications done by classification methods. It is one $n \times n$ matrix. 32 different plant species are used for the methods. Thus, n equals 32.

Precision and recall values are calculated for each label of the used methods by using confusion matrix information. Generally, precision is a measure of accuracy provided that a specific class has been predicted. Recall is a measure of the ability of a prediction model to select instances of a certain class from a dataset. The following equations, equation 3 and equation 4, are used to obtain precision and recall values from confusion matrixes.

$$Precision_i = \frac{M_{ii}}{\sum_j M_{ji}} \quad (3)$$

$$Recall_i = \frac{M_{ii}}{\sum_j M_{ij}} \quad (4)$$

Where i is i^{th} row of confusion matrix and j is the j^{th} column of it.

In both Figure 5 and Figure 6, HARRIS-SURF method has the minimum values. It is predictable as it has the least accuracy percentage between three methods. Variation of SURF method values is less than other methods in both figures. Precision is a measure of result relevancy, while recall is a measure of how many truly relevant results are returned. The SURF method has the best performance due to the figures. FAST-SURF method has the second rank between these three methods. Its performance is in the middle. Due to the measurements, the sequence is SURF, FAST-SURF, and HARRIS-SURF.

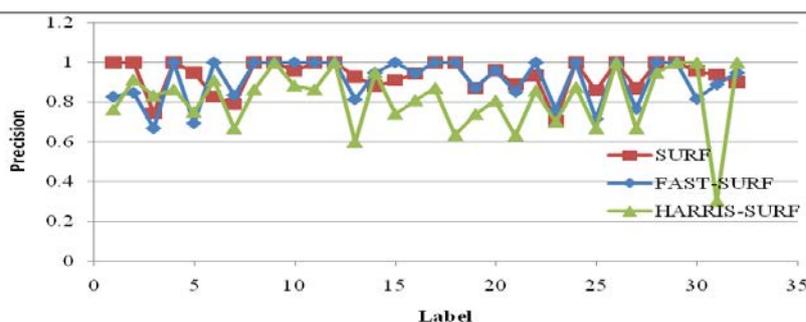


Figure 5: Precision measurement for SURF, FAST-SURF, HARRIS-SURF methods.

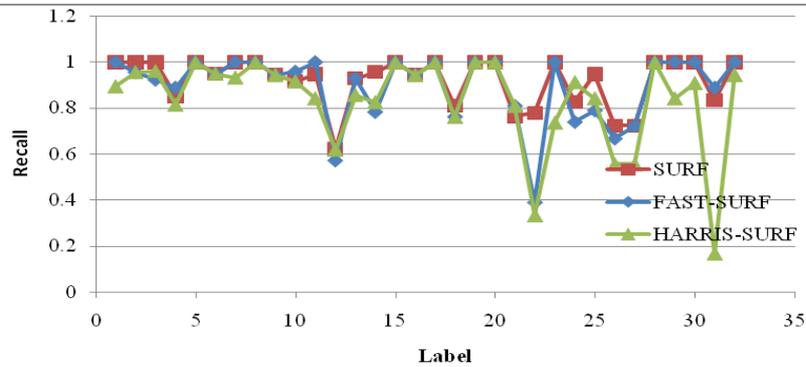


Figure 6: Recall measurement for SURF, FAST-SURF, HARRIS-SURF methods.

Another concept to compare the methods is surrounded area. Higher area under the curve represents both higher precision and higher recall. A lower false positive rate relates to high precision and a lower false negative rate relates to high recall. Due to these concepts in Figure 5 and Figure 6, SURF method has better performance of the method because higher areas belong to it. 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 is shown for each method.

For SURF method, the minimum value of recall is 0.619048 and the minimum value of precision is more than 0.703704. Due to experiment, variation of values in SURF method is less than the SIFT method used in [19]. In Figure 7, these variations are shown.

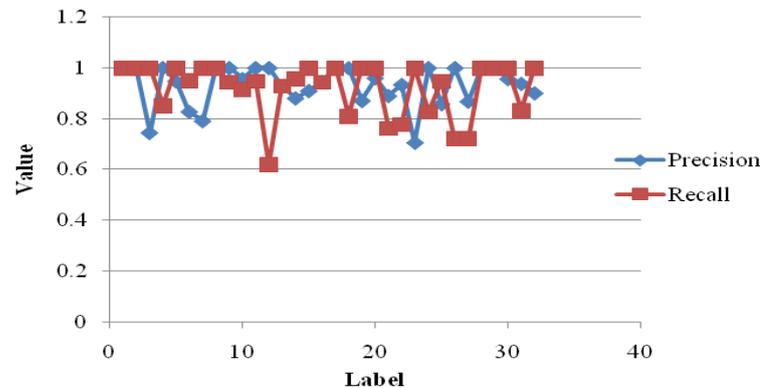


Figure 7: Precision and recall values for SURF method.

Also, more labels have values in $[0.9, 1]$, which is the highest possible interval.

In Figure 8, precision and recall values are shown for FAST-SURF method. In this method, the minimum values of precision and recall are the same and equal to 0.666667. In comparison to SURF method, there are more variations in obtained results of precision and recall. After investigation of minimum value of recall, it is found that the value is less than 0.4 and equals 0.3333. In comparison to the used FAST-SIFT method in [19], this combined method has better performance.

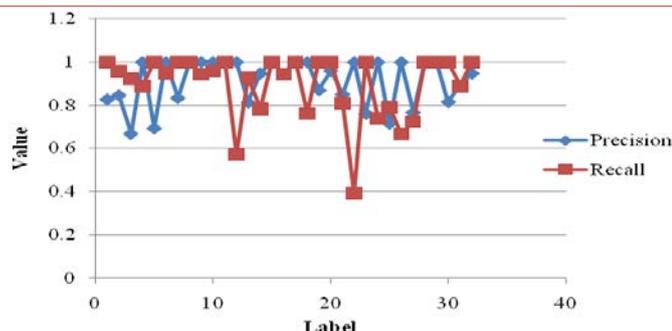


Figure 8: Precision and recall values for FAST-SURF method.

The Figure 9 shows precision and recall of HARRIS-SURF method. The minimum value of precision is 0.3 while the minimum value of recall is 0.166667. Difference between maximum and minimum is high in this method for both precision and recall values. Larger intervals are covered with HARRIS-SURF method for obtained precision and recall values. The reason is completely obvious, because the accuracy of this method is lower than other methods, SURF and FAST-SURF methods.

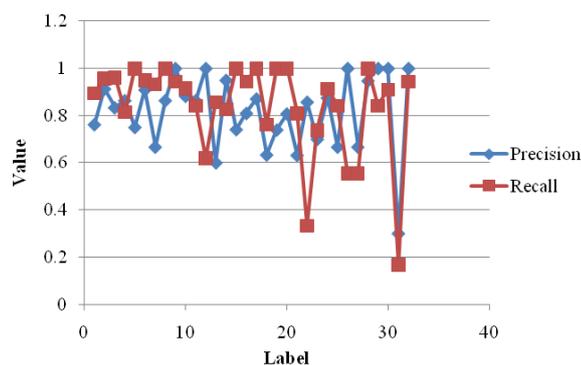


Figure 9: Precision and recall values for HARRIS-SURF method.

5 Conclusion

In this paper, three methods are taken into consideration for plant recognition and identification on Flavia dataset. SURF method and two combined methods, HARRIS-SURF and FAST-SURF, are taken into consideration for plant recognition and classification. Accuracy measurement and efficiency of each method are described. 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 SURF method has the best performance between proposed methods. In comparison with used methods in [19], SURF and FAST-SURF have better performance and accuracy. Improvement can be done for next step and obtain higher accuracy.

REFERENCES

- [1]. Allen I. White, *The History of the Washington State University College of Pharmacy 1891-1991*. 1996. pp. 63-65.
- [2]. T. K. C. Im, H. Nishida, *Recognizing plant species by leaf shapes-a case study of the acer family*. Proceedings of IEEE International Conference on Pattern recognition, 1998. pp. 1171.

- [3]. Z. Wang, Z. Chi, and D. Feng, *Shape based leaf image retrieval*, in Vision. Image and Signal Processing, IEE Proceedings, 2003. 150(1): pp. 34–43.
- [4]. J. Du, D. Huang, X. Wang, and X. Gu, *Computer-aided plant species identification (CAPSI) based on leaf shape matching technique*. Transactions of the Institute of Measurement and Control, 2006. 28 (3): pp. 275-284.
- [5]. S. G. Wu, F. S. Bao, E. Y. Xu, Y.-X. Wang, Y.-F. Chang, and Q.-L. Xiang, *A leaf recognition algorithm for plant classification using probabilistic neural network*. IEEE International Symposium on Signal Processing and Information Technology, IEEE explore library, 2007. pp. 11–16.
- [6]. Y. Li, Q. Zhu, Y. Cao, and C. Wang, *A leaf vein extraction method based on snakes technique*. Proceedings of IEEE International Conference on Neural Networks and Brain, 2005.
- [7]. S. Belongie, J. Malik, and J. Puzicha, *Shape Matching and Object Recognition Using Shape Contexts*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002. 24(4): pp. 509–522.
- [8]. H. Ling and D. W. Jacobs, *Shape Classification Using the Inner-distance*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007. 29(2): pp. 286–299.
- [9]. Herbert Bay, Tinne Tuytelaars and Luc Van Gool, *SURF: Speeded Up Robust Features*. Computer Vision – ECCV 2006, Lecture Notes in Computer Science, 2006. 3951: pp. 404-417.
- [10]. P. M. Panchal, S. R. Panchal, S. K. Shah, *A Comparison of SIFT and SURF*. International Journal of Innovative Research in Computer and Communication Engineering, 2013. 1(2).
- [11]. Lowe, D. G., *Object Recognition from Local Scale-Invariant Features*. Proc. Of the International Conference on Computer Vision, 1999. pp. 1150—1157.
- [12]. G. Csurka, C. Dance, L. Fan, J. Williamowski, and C. Bray. *Visual categorization with bags of keypoints*, in. ECCV'04 workshop on Statistical Learning in Computer Vision, 2004. pp. 59–74.
- [13]. Xin Chen, Xiaohua Hu¹, Xiajiong Shen, *Spatial Weighting of Bag-of-Visual-Words*. Proceedings of the 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, 2009. pp. 867-874.
- [14]. Yi Yang, Shawn Newsam, *Bag-of-visual-words and spatial extensions for land-use classification*. Proceedings of SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2010. pp. 270-279.
- [15]. Michael Villamizar, Jorge Scandaliaris, Alberto Sanfeliu, and Juan Andrade-Cetto, *Combining Color-based Invariant Gradient Detector with HoG Descriptors for Robust Image Detection in Scenes under Cast Shadows*. IEEE International Conference on Robotics and Automation, 2009. pp. 1997-2002.

- [16]. Jun Yang, Yu-gag Jiang, *Evaluating bag-of-Visual-words representations in scene classification*. Proceedings of the International Workshop on Multimedia Information Retrieval, 2007. pp. 197-206.
- [17]. David Picard, Nicolas Thome and Matthieu Cord, *An Efficient System for Combining Complementary Kernels in Complex Visual Categorization Tasks*. Proceedings of IEEE International Conference of Image Processing (ICIP), 2010. pp. 3877-3880.
- [18]. C.-F. Tsai, *Bag-Of-Words Representation in Image Annotation: A Review*. International Scholarly Research Network, 2012. pp. 1-19.
- [19]. Masoud Fathi Kazerouni, Jens Schlemper, and Klaus-Dieter Kuhnert, *Comparison of Modern Description Methods for the Recognition of 32 Plant Species*. Signal & Image Processing: An International Journal (SIPIJ), 2015.
- [20]. H. Laga, S. Kurtek, A. Srivastava, M. Golzarian, and S. Miklavcic, *Ariemannian elastic metric for shape-based plant leaf classification*. Digital Image Computing: Techniques and Applications, 2012.
- [21]. K. Mikolajczyk, and C. Schmid, *Scale and Affine Invariant Interest Point Detectors*. International Journal of Computer Vision, 2004. 60(1): pp. 63-86.
- [22]. A. Bosch, X. Munoz, and R. Marti, *Which is the best way to organize/classify images by content?*. Image and Vision Computing, 2007. 25(6): pp. 778–791.
- [23]. K. Mikolajczyk, B. Leibe, and B. Schiele, *Local features for object class recognition*. Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05), 2005. pp. 1792–1799.
- [24]. T. Tuytelaars and K. Mikolajczyk, *Local invariant feature detectors: a survey*, Foundations and Trends. Computer Graphics and Vision, 2007. 3(3): pp. 177–280.
- [25]. C. Harris, M. Stephens, *A Combined Corner and Edge Detector*, in Proceedings of the Fourth Alvey Vision Conference, 1988. pp. 147-151.
- [26]. E. Rosten, T. Drummond, *Machine Learning for High-Speed Corner Detection*, European Conference on Computer Vision, 2006. 1: pp. 430–443.
- [27]. V. Vapnik, *The natural of statistical learning theory*, Springer-Verleg, New York, USA, 1995
- [28]. S. M. Smith and J. M. Brady, *SUSAN — A new approach to low level image processing*, International Journal of Computer Vision, 1997. 23(34): pp. 45–78.