Automatic Plant Recognition System for Challenging Natural Plant Species

Masoud Fathi Kazerouni
Institute of Real-time Learning Systems
Hölderlinstr. 3
Germany D-57076, Siegen, NRW
masoud.fathi@uni-siegen.de

Jens Schlemper
Institute of Real-time Learning Systems
Hölderlinstr. 3
Germany D-57076, Siegen, NRW
schlemper@fb12.uni-siegen.de

Klaus-Dieter Kuhnert
Institute of Real-time Learning Systems
Hölderlinstr. 3
Germany D-57076, Siegen, NRW
kuhnert@fb12.uni-siegen.de

ABSTRACT

Photosynthesis is one of turning points to shape the world. Plants use this process to convert light energy into chemical energy. Some of the early microorganisms evolved a way to use the energy from sunlight to make sugar out of simpler molecules, but unlike green plants today, the first photosynthesizing organisms did not release oxygen as waste product, so there was no oxygen in the air. Plants are very busy factories and leaves are the main place for production. A useful plant recognition system is capable of identification of different species in natural environment. In natural environment, plants and leaves grow in different regions and climates. During day, variation of light intensity can be considered as an important factor. Thus, recognition of species in different conditions is a real need as plants are ubiquitous in human life. A dataset of natural images has been utilized. The dataset contains four different plant species of Siegerland, Germany. Modern combined description algorithms, SURF, FAST-SURF, and HARRIS-SURF, have been carried out to implement a reliable system for plants species recognition and classification in natural environment. One of well known methods in machine learning community, Support Vector Machine, has been applied in the implemented systems. All steps of system's implementation are described in related sections. The highest obtained accuracy belongs to the implemented system by means of SURF algorithm and equals to 93.9575.

Keywords

Component, SURF, combination, FAST, HARRIS, FAST-SURF, HARRIS-SURF, feature extraction, feature detection, natural images, natural plant species recognition, weather condition, light intensity

1. INTRODUCTION

Ever since early man rubbed two sticks together to make fire, plants have played a vital role without any doubt in the history of mankind. In past years, some plants have been become extinct and risk of extinction always exists for plants. Collection of their information is essential to have a complete database of different plants species [1, 2, 3, 4].

Developing systems that assist botanists and professionals to recognize and identify types of plants species is really necessary in modern life. Plant species recognition relies on computational methods to extract discriminative features from images like other image recognition tasks. Tendency to have automatic systems has been increased recently. Thus a set of techniques that learn features automatically has priority. It is a goal to transform raw data to correct representation which can be effectively exploited in machine learning and pattern recognition tasks. Two main advantages of feature learning are automatic analysis of images and efficient use of features in image classification.

Real-world data is usually complex, redundant, variable, and even noisy. Adaption of present strategies is very important to automate and generalize plant recognition system.

To build a useful system, different aspects or components of natural environment should be taken into consideration. More precisely, climate, weather, and wilderness can affect on performance of a plant recognition system. Therefore, weather condition is one of added factors. Different types of natural light can produce a wide variety of appearances when an image is taken from a plant, even though the light source, sun, is the same. Position and distance of leaves from camera are different, and point of view has influence on the image too. Time of day is another factor which can be considered in developing a helpful and accurate system. The system should be independent from the camera which is used to take plants images. Consideration of these continuum factors contributes to have a reliable and general system for various agricultural and medicine applications. The implemented system should tackle all challenges according to mentioned factors.

The prepared and used dataset in this work is more challenging than other present datasets as it does not contain scanned or pseudo-scanned images [5], [6]. It is one of the most realistic references of the current state of art in this area. Figure 1 shows some sample images of the dataset which belong to one plant species.

Figure 1. Four different sample images of Cornus.

Previous works have some disadvantages. Some of them [7, 8] are only applicable to certain species and they cannot be applied on other species, therefore they lack the factor of being useful for general applications. In [9], the implemented system is called a semi-automatic system, experiment is performed on 1000 images, and the obtained accuracy is 85%. Being semi-automatic is one of its weaknesses. In [10] and [11], modern combined detection and description methods have been used for classification of 32 different plant species. The used dataset in [10] and [11] is not natural, although it is a common dataset.

In this work three combined methods, SURF [12], FAST-SURF [11], and HARRIS-SURF [11], has been utilized to extract features and classify images with machine learning methods automatically. The main purpose of this work is to consider real parameters and natural factors and invent a system according to them. This paper presents some contributions. First, the proposed systems are full automatic. Second, combined methods have been used to do modern detection and description phases. Third, natural images have been utilized for the whole work. Fourth, some environmental parameters such as light intensity, and weather condition are not kept fixed.

The rest of this paper is organized as follows: section 2 presents the plant identification task; section 3 describes the details of proposed approaches; section 4 presents conducted experiments and obtained results. Finally, conclusions are drawn in section 5.

2. GENERAL REVIEW

Plant recognition task is a huge problem that had been escaped into neglect for years. In recent years, more attention has been paid to develop automatic plant recognition systems which have been demanded in different fields of scientific activities and industry. Development of such a system is important for exploration and protection of plant species to have essential information and relationship between plants.

In developing such a useful system, different parameters should be considered in all designing and implementation steps. Some of parameters are compatibility of the system with different environmental and illumination conditions, weather, distance, and etc. In order to achieve the desired goals and have a system which has generality, the first step is determination of an appropriate dataset. The dataset must include natural images to be completely close to real world. Natural images are photographs of the surrounding environment where people live. Rich covariance structure is one of natural images' properties. Characteristics of information related to redundancy are very important in classification tasks. In natural images, background of each image is different from the other one. Although it makes the recognition procedure harder, it helps to have a general dataset. According to mentioned facts, selection of an efficient method to extract features is vital and affects other classification steps. To have sufficient information, modern and engineered methods can be applied to extract effective features.

Input raw images to algorithms and systems are too large to be processed. In feature detection step, significant locations of image are produced as output. For instance, corner detectors find locations of corners in one image. To get useful information of detected parts, encoding interesting information into a series of numbers will be done by feature description methods. In computer vision field, there are various feature detection methods such as HARRIS [13] and FAST [14] which can be applied to relax the detection step and complexity of initial images. In [8] and [15], SIFT, FAST, and HARRIS methods have been utilized to construct combined feature detection and description methods. In [10], one common component of proposed combined methods is SIFT algorithm, and the system can recognize 32 different plant species and the highest obtained accuracy is 89.3519%. SURF method also can be used as a high performance scale and rotation-invariant interest point detector and descriptor [11]. This method will be used in this work to implement an automatic plant recognition system.

In order to compensate disadvantages of current systems [10, 15], another modern algorithm, SURF, is applied to implement the desired automatic system. Its speed is higher than SIFT algorithm and has good performance. Application concerns are very important for choosing an algorithm. Due to execution time of the system by SIFT in [15] for plant recognition, it is decided to change the system by using SURF algorithm as the core of methods. Multi resolution pyramid technique is utilized in SURF algorithm to make a copy of the original image with Pyramidal Gaussian or Laplacian Pyramid shape to obtain an image with the same size but with reduced bandwidth. Thus a special blurring effect on the original image, called Scale-Space, is achieved [16]. This technique ensures that the points of interest are scale invariant [16]. SURF algorithm shows good behavior in view of slight perspective changes. In [11], SURF method has been used to distinguish 32 different plant species, where the used dataset is artificial and the images have been recorded against a white background, and the obtained accuracy is 92.28395%. This accuracy is higher than other systems which have been proposed in [10] and [17]. In order to classify plant species, Probabilistic Neural Network (PNN) has been utilized in [17] and the obtained accuracy of this semi-automated system is 91.41%. Being semi-automatic is one its disadvantages and, it is tested on only artificial images.

By using Bag of Visual Words, an image can be represented in one feature vector. However this technique is originally often used in natural language processing, it can be applied to images and a word can be considered as amount of local features. When images are expressed by vectors, it is possible to use support vector machines (SVMs) [18] for classification step.

The root of SVMs is Statistical Learning Theory and they have been mostly used in classification problems [19] or regression challenges [20]. They are robust, accurate, and effective even when a small training dataset is used. Also, they have been adopted to handle multiple classification tasks. In [21], SVM has been applied to classify leaves efficiently where the input vector of SVM is 5 variables which have been composed of 12 leaf features. In [15], SVMs have been applied in the proposed systems to do classification of natural images of plant species.

The final step is the test part. This part is performed over 336 images of the dataset. All experiments have been performed and investigated on these natural images for 4 different plant species.

3. Plant Recognition System Investigation

Automatic plant recognition systems are required to operate at a high level of reliability and accuracy in natural environment and real world. The system explanation is simplified by dividing the system into 4 different steps. The first step is the preprocessing part, which deals with conversion of RGB image. The feature detection and description parts are considered as the second step. The third step is the training part of the system in detail. The forth step consists of the test procedure of implemented systems.

3.1.1 Preprocessing

Inputs are natural images in RGB model in the prepared dataset. There are different conversion methods to convert images to grayscale format. This step is performed as the preprocessing part of the proposed system. As dataset includes natural images, each image has different numbers of objects, background, and etc. The following equation, Equation 1, has been used for conversion from RGB format to grayscale format [22].

$$Y = 0.299\,R + 0.587\,G + 0.114\,B \quad \text{(Equation 1)}$$

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

3.1.2 Detection and Extraction of Features

The next step is detection and extraction of features. Undoubtedly human eye can extract all information of raw image, but all information is not useful for computer algorithms. In real world, changes of illumination, viewpoints, and scales are irrefutable, therefore local representations have priority to be used in proposed system. The important point is that global representations are not invariant to transformations and they are sensitive to changes. On other words, local methods represent images based on some salient regions, which remain invariant to viewpoints and illumination changes.

Due to superior performance of local features [23], these types of methods have been selected for the current systems. Also local structures are very helpful for object recognition and classification applications. Meanwhile they lead to achieve high accuracy. Due to large numbers of local features, the amount of memory increases in comparison to other methods, and it is one disadvantage of local representation. The solution is to aggregate local image descriptors into compact vector representation [24].

A reliable method in presence of changes and deformations is essential, and it can affect the whole procedure. Three various categories of feature detectors which are single scale detectors, multi scale detectors, and affine invariant detectors can be considered. In 1977, Hans P. Moravec defined the concept of interest points as distinct regions [25]. He was interested in finding distinct regions that could be used to register consecutive image frames [25]. This method is not invariant to rotation, and also it has a low repeatability rate. After 11 years, Harris and Stephens proposed a new method, Harris detector, to improve limitations of the Moravec’s algorithm. Harris detector is a good combined corner and edge detector which has roughly acceptable detection results and repetition rate. It is based on intensity variation over all directions. Corners are the regions in the image with large variation in intensity in all directions. Auto-correlation matrix is a very popular mathematical technique which is utilized in features detection methods. One 2×2 symmetric auto-correlation matrix is used in Harris method as follows:

$$M(x,y) = \sum_{u,v} w(u,v) * \begin{bmatrix} I_x^2(x,y) & I_xI_y(x,y) \\ I_xI_y(x,y) & I_y^2(x,y) \end{bmatrix}$$

(Equation 2)

where I_x is local image derivative in the *x* direction, nd I_y is local image derivative in the *y* direction, and $w(u, v)$ indicates a weighting window, rectangular or circular window, over the area (u,v). Moreover, window function gives weights to pixels.

To apply a circular window, a Gaussian one should be used. In this case, the response is isotropic, and the values will be weighted more heavily near the center. Finding interest points will be done by computation of eigenvalues of the mentioned matrix for each pixel. When both eigenvaules are large, it means that it is the location of a corner. In order to measure corner $C(x,y)$ for each pixel (x,y), the following equation will be used:

$$C(x, y) = det(M) - K(trace(M))^2 \quad \text{(Equation 3)}$$

where

$det(M) = \lambda_1 * \lambda_2$, and $trace(M) = \lambda_1 + \lambda_2$ (6) (Equation 4)

K is used as an adjustment parameter and the eigenvalues of the auto-correlation matrix are λ_1, λ_2. Harris proposed to combine the eigenvalues in a single measure instead of two measures. Furthermore, the obtained eigenvalues make decision on status of a region. When λ_1, and λ_2 are small, $|C|$ is small too, then the region is flat, and the windowed region has constant intensity approximately. For example, the region is flat when there is little change in C in any direction. When one eigenvalue is high and the other one is low ($\lambda_2 >> \lambda_1$ or vice versa), C will be less than zero and the region is edge. In other words, the local auto correlation function is ridge shaped and little change in C has been caused by local shifts in one direction along the ridge, in addition to, significant change happens in the orthogonal direction. The last condition happens when λ_1 and λ_2 are large, so the local auto correlation function is peaked sharply and shifts in any direction will cause an increase. In this condition, C is large and the region is one corner. Harris has been used as one of detection algorithms.

Another algorithm has been applied to detect keypoints as it is an important to identify correspondence of keypoints in images. A FAST algorithm is very attractive for this purpose to get keypoints in a high speed level. It is an efficient corner detector based on comparing pixels intensities. In this algorithm, a circle of 16 pixels surrounding the central pixel has been considered to identify corners. In classification part, at least 12 continuous pixels must be darker or lighter than the central pixel, and then the central pixel will be corner. When central the pixel is classified as a corner pixel, it is not necessary to test all 16 pixels in cases of a non-corner pixel. Thus the algorithm is quick and applicable when high speed is needed, but it detects a large number of corners which is a disadvantage. This disadvantage arises from the basis of the algorithm as it is based on intensity information of 16 surrounding pixels. Furthermore, natural images are complex and a large number of corners will be detected by FAST. One pixel, *p*, is selected in the image. To know the status of this pixel, whether it is a corner or not, the algorithm works as below.

On threshold is set due to the selected pixel's intensity value. It is 20% of this pixel, and is considered as T. A circular area, which is consisted of 16 pixels, is considered as a search region to find out the pixel's status. 12 pixels out of these 16 pixels should be higher or lower than T (value of I_p). To have a fast algorithm, intensity of pixels 1, 5, 9, 13 are compared with Ip. Three of these four pixels should satisfy the threshold condition, so the selected pixel will exist and remain. When at least three of the four pixel values - I_1 ,I_5 ,I_9, I_{13} are not above or below $I_P + T$, p is not a corner. In this case, the pixel p is rejected. All 16 pixels are checked when at least three of the considered pixels are above or below $Ip + T$. In this case, if 12 contiguous pixels have the required condition, the selected pixel is a corner. In this algorithm all pixels will be tested and corners will be found finally.

The last used algorithm is SURF, which is applied for both feature detection and extraction. The algorithm speeds up the SIFT algorithm without scarifying the quality of detected points. Thus this method is widely used in different computer vision applications due to its efficiency, distinctiveness, and robustness in invariant feature localization. Also it has been applied to extract features as a component of the used combined methods. In SURF algorithm, an intermediate image representation, called image integral [26], is used to increase the calculation speed of the algorithm. The integral image is obtained by computation of an input image. The input image is *I*, and integral image is *IM* where a point is *(x,y)*.

$$IM(x,y) = \sum_{i=0}^{i\le x} \sum_{j=0}^{j\le y} I(x,y) \quad \text{(Equation 5)}$$

When the integral image is used, calculating the area of an upright rectangular will lead to reduction of four operations. In addition to, change of size does not affect computation time and the algorithm is useful, even though large areas are required.

SURF entails computation of Hessian matrix and its detector is basically based on the determinant of this matrix. A 2-dimensional Hessian matrix consists of a *2*2* matrix containing the second order partials derivatives of a scalar valued function (image pixel intensities) as shown in below:

$$H(f(x,y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \qquad \text{(Equation 6)}$$

It is a symmetric matrix and its determinant is the product of eigenvalues.

Calculation of derivatives is performed by convolution with a suitable kernel. The determinant can be calculated in different scales. Gaussians are optimal for scale space considerations. SURF approximates LoG with box filter. A parallel procedure is also possible due to two usages, box filters and integral images. The purpose is to increase computational efficiency. Integral images help to do fast computation of box convolutions and have a fast way to compute intensities for any rectangle within the image, which is independent of rectangle size. Also, computation time is not sensitive to the size of filter. A scale space is divided into octaves. These octaves show a series of filter response maps obtained by convolving the same input with a filter of increasing size [27].

The construction of scale space begins with a *9*9* filter. It calculates blob response of the image for the smallest scale. After that, the sizes of filters increase to continue the procedure, *15*15*, *21*21*, *27*27*, and etc. Blob response is shown at location $x = (x, y, \sigma)$ as the following:

$$det\,(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2 \qquad \text{(Equation 7)}$$

When the used filter is a *9*9* matrix, σ equals to 1.2. In general, the following equation exists:

$$\sigma = (current\ filtersize/base\ filter\ size)*(base\ filter\ scale) \qquad \text{(Equation 8)}$$

where base filter size is 9 and base filter scale is 1.2.

An important step is to localize and find major keypoints in scale space. To achieve this goal, a non-maximum suppression in a *3*3*3* neighborhood is applied. The maxima of the determinant of the Hessian matrix are then interpolated in scale and image space with the method proposed by Brown et al. [27, 28]. One considerable point is the difference in scale between the first layers of every octave which is large, thus scale space interpolation is an important issue.

The next part is feature description, which should be robust and unique for a feature. Other points are the direct impact computational complexity, robustness, and accuracy. In description phase, the bases are on Haar wavelet responses in horizontal and vertical directions, x and y. The integral images will be used to do efficient calculations at any scale. Finding orientation of interest points contributes to have rotational invariance algorithm. Gaussian weights are applied to the interest point obtain robustness against deformations and translations. SURF provides a functionality which is called Upright-SURF or U-SURF and it contributes to robustness up to $\pm 15^{\circ}$ [27]. This version of SURF improves the speed of algorithm as one of advantages.

An interest area is defined by a window size of *20s*20s*. This area is divided into *4*4* square regions as subareas, and they contribute to keep spatial information. Then, Haar wavelet responses are computed for each subarea in x and y directions. A vector is created after this procedure. In the following, the formed vector is shown:

$$v = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|) \qquad \text{(Equation 9)}$$

In general, there are two different dimensions for SURF feature descriptor, 64 and 128. 64 dimension version has higher speed and 128 dimension version provides better distinctiveness of features. It can be considered as another functionality of the SURF algorithm. Another improvement of this algorithm helps in matching stage and speed up it. It is sign of Laplacian and the purpose is to distinguish bright blobs on dark backgrounds and vice versa:

$$\nabla^2 L = tr(H) = L_{xx}(x,\sigma) + L_{yy}(y,\sigma) \qquad \text{(Equation 10)}$$

As Laplacian is the trace of Hessian matrix, the values have been calculated for determinant of the Hessian matrix before. Also, it is possible to use the sign of Laplacian to have faster matching, and it does not have any impact on performance of description and other stages. This algorithm, SURF, is used as description component of FAST-SURF and HARRIS-SURF methods. The next step of the system is training model, which includes two main parts, BoW and using SVMs for training.

3.1.3 Training

A useful and successful approach for recognition tasks is to quantize local visual features and then apply BoW method. Then a classifier, such as SVM, is applied to do classification task. By using this method, each image will be described as collection of words. In this step, obtained local descriptor will be proceeded to build a codebook and its final output will be utilized in next step. Also, this method has been widely applied for image representation [29], [30], and [31]. In [10], [11], and [15], BoW technique has been selected to classify different plant species, because it is unaffected by position and orientation of object in image and it applies fixed length vector irrespective of number of detections. The used dataset contains natural images in this system. For instance, illumination, light intensity, time, scale, and distance are not the same in all images and they vary. Due to the characteristics of the dataset, this method contributes to represent and use images in right way and format, hence it is the right choice.

In order to create a vocabulary of visual words, quantization and grouping of local descriptors are essential. Each group will be one specific word of the vocabulary (dictionary). Therefore, K-means clustering [32] has been applied in this part due to its efficiency in clustering applications. Also it is a usual quantization approach due to its simplicity and high convergence speed. It is possible to reduce the size of feature space and compute histograms related to visual words. A finite number of clusters will be available to form a visual vocabulary. K-means uses Euclidean distance. This method finds the positions y_i, $i= 1, 2, 3, ..., k$, of the clusters that minimize the following equation, the square of the distance from the obtained feature descriptors to the cluster.

$$\arg\min_{c}\sum_{i=1}^{k}\sum_{x\in c_i} d(x,\mu_i)^2 = \arg\min_{c}\sum_{i=1}^{k}\sum_{x\in c_i} \| x-\mu_i \|_2^2$$

(Equation. 11)

where k is the number of clusters, and μ_i is the centroid of all points in c_j.

Euclidean distance is a base of this step, and local descriptors are assigned to the nearest word. As the distance between camera and plants varies, constructed vocabulary is different for each distance.

3.1.4 SVM Classification

A classifier has been designed and built in this challenging step. It has been proven that SVMs are useful tools in real world applications based on statistical learning theory. In image recognition, supervised machine learning models have been considered as efficient methods in many cases. SVM involves in finding separating optimal hyper-plane in higher dimensions efficiently, which maximizes the margin of the training data. When number of dimensions is greater than number of samples, SVM stays efficient and it is one of its benefits in addition to efficient memory. In [10], [11], [15], and [33], SVMs have been used as a part of implemented systems and the results show that SVMs are effective and efficient. The key features of SVMs are the use of kernels, the absence of local minima, the sparseness of the solution and the capacity control obtained by optimizing the margin. Therefore, SVMs have been applied in three implemented systems.

4. RESULTS AND DISCUSSIONS

Investigation of each implemented system has been performed in this section individually. A new plant dataset which consists of 4 classes, collected at the Learning Real-time System Institute, Siegen, Germany, is employed in the experiments. These images are sampled in the campus of Hölderlin, University of Siegen, Germany. Evaluation experiments have been performed to test effectiveness of the proposed systems in different aspects. The dataset includes 1000 images, and divided into two sub-datasets, training and test datasets, and corresponding numbers of sub-datasets are 664 and 336 images. Table 1 shows the number of images for each sub-dataset and relevant distances. Images' variations of the dataset are large in the prepared dataset as it has been pointed out before.

Dataset	25 cm	50 cm	75 cm	100 cm	150 cm	200 cm
Number of images for Training	160	160	160	160	12	12
Number of images for Test	80	80	80	80	8	8

Table 1. Number of images in different distances

The hardware used for the experiments is a personal computer (PC) equipped with an Intel® Core™ i7-4790K, CPU @ 4.00 GHz, and installed memory (RAM) 16.0 GB.

The first experiment is finding accuracy of the proposed systems in each distance. In order to obtain accuracy, total number of correct predictions is divided by total number of predictions [34], then it is multiplied by 100 and percentage of accuracy is calculated. While distance between camera and plants is 25 cm, the results show that implemented methods which are SURF method and HARRIS-SURF method, have the highest accuracy between all of them. Their accuracy is equal to 95%, and FAST-SURF has 93.75 percentage of accuracy.

Applied Method	Correct Predictions	Wrong Predictions	Percentage of Accuracy
SURF	76	4	95.0000
FAST-SURF	75	5	93.7500
HARRIS-SURF	76	4	95.0000

Table 2. Accuracy of classification (Distance 25 cm- K= 1000)

The accuracy of SURF method remains constant, when the distance increases to 50 cm. In this case, decrease of accuracy is high by using HARRIS-SURF method in comparison to previous distance and it has the lowest accuracy between three implemented methods. The accuracy of the system with FAST-SURF is 91.25% and increase of distance from 25 cm to 50 cm has less effect on it.

Applied Method	Correct Predictions	Wrong Predictions	Percentage of Accuracy
SURF	76	4	95.0000
FAST-SURF	73	7	91.2500
HARRIS-SURF	69	11	86.2500

Table 3. Accuracy of classification (Distance 50 cm- K= 1000)

Increase of distance affects on accuracy of implemented systems. Implemented SURF system has the highest accuracy in 75 cm, where two other systems have the same accuracy and it is 85%.

Applied Method	Correct Predictions	Wrong Predictions	Percentage of Accuracy
SURF	72	8	90.0000
FAST-SURF	68	12	85.0000
HARRIS-SURF	68	12	85.0000

Table 4. Accuracy of classification (Distance 75 cm- K= 1000)

The last accuracy measurement is done for 100 cm, 150 cm, and 200 cm distances. In this case, the system with SURF method has highest accuracy, 95.83%, and accuracy of the other systems is equal to 93.75%. The results are shown in Table 5.

Used Method	Correct Predictions	Wrong Predictions	Percentage of Accuracy
SURF	92	4	95.8300
FAST-SURF	90	6	93.7500
HARRIS-SURF	90	6	93.7500

Table 5. Accuracy of classification (Distance 100 cm, 150 cm, 200 cm- K= 1000)

Table 6 includes the information of total number of correct and wrong predictions for three implemented systems. Results of all distances have been used to achieve the following table. Highest accuracy has been achieved by SURF method, where correct predictions are 316 out of 336 tested images.

Applied Method	Correct Predictions	Wrong Predictions	Percentage of Accuracy
SURF	316	20	93.9575
FAST-SURF	306	30	90.9375
HARRIS-SURF	303	33	90.0000

Table 6. Accuracy of classification (K= 1000)

In order to understand, investigate, and describe performance and quality of an automatic classification system, confusion matrix is often used, which is an easy understanding table. It is a popular tool for performance evaluation and predictive capability of a classification system. Additionally, nature of the classification errors can be identified and compared. Each table represents one square matrix 4-by-4 for each method and distance. The confusion matrixes are shown in Table 7, Table 8, and Table 9.

SURF Method	Hydrangea	Amelanchier Canadensis	Acer Pseudoplatanus	Cornus
Hydrangea	78	0	5	1
Amelanchier Canadensis	5	76	1	2
Acer Pseudoplatanus	1	0	82	1
Cornus	2	2	0	80

Table 7. Confusion matrix- SURF

FAST-SURF Method	Hydrangea	Amelanchier Canadensis	Acer Pseudoplatanus	Cornus
Hydrangea	74	0	8	2
Amelanchier Canadensis	4	73	2	5
Acer Pseudoplatanus	1	1	81	1
Cornus	2	4	0	78

Table 8. Confusion matrix- FAST-SURF

HARRIS-SURF Method	Hydrangea	Amelanchier Canadensis	Acer Pseudoplatanus	Cornus
Hydrangea	72	0	7	5
Amelanchier Canadensis	5	72	1	6
Acer Pseudoplatanus	1	1	81	1
Cornus	3	1	1	79

Table 9. Confusion matrix- HARRIS-SURF

Precision and recall results have been derived from confusion matrix. Obtained results of precision and recall calculations show that variations of the HARRIS-SURF method's results are less than two other methods in 25 cm distance. In Figure 2, precision and recall calculations are shown for this distance. FAST-SURF method has the lowest area under its curve in precision case. It could be expected as the lowest accuracy also belongs to this method. Comparison of results can be done by investigation of area under the curves. A high area under one 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.

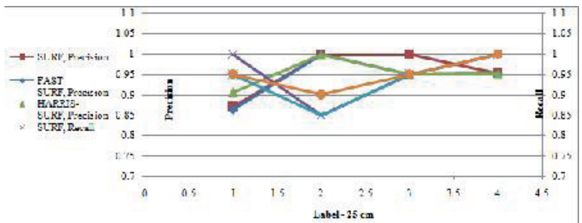

Figure 2. Precision and recall- 25 cm.

In Figure 3, precision and recall measurements are shown. SURF method has the highest areas under precision and recall curves between three used methods. The minimum values of precision and recall measurements belong to HARRIS-SURF method.

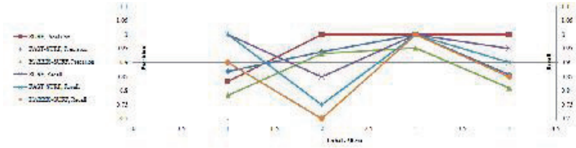

Figure 3. Precision and recall- 50 cm.

Increase of distance affects on performance of applied methods in proposed automatic classification system. As a result, measures of precision and recall calculations will be changed. In distance 75 cm, SURF method has the highest accuracy. Hence, surrounded areas under precision and recall curves are more than other two methods. In Figure 4 precision and recall results of three used methods are shown. The shape of FAST-SURF method's curve is more similar to SURF method's curve than HARRIS-SURF method's curve in precision measurements.

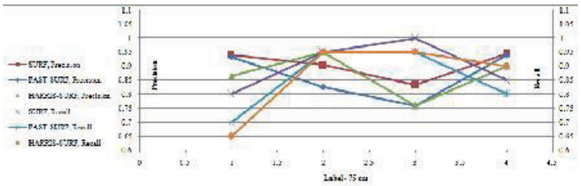

Figure 4. Precision and recall- 75 cm.

When the distance increases from 75 cm to 100 cm, 150 cm, and 200 cm, recall and precision results have been changed. The SURF method has the highest accuracy, and the recall values have been in a range, where the range is [0.91667, 1]. Precision results' range is [0.92308, 1]. These two different ranges show that the obtained values of precision and recall measurements are large and prove that system can return many correct labeled results. In this special case, obtained results of HARRIS-SIFT method are comparable to the results of FAST-SIFT method, because they have the same accuracies. The difference between accuracy of SURF method and HARRIS-SURF method and FAST-SURF method is 2% and it is small, therefore they can be applied according to desired usages and applications. Figure 5 shows the mentioned points.

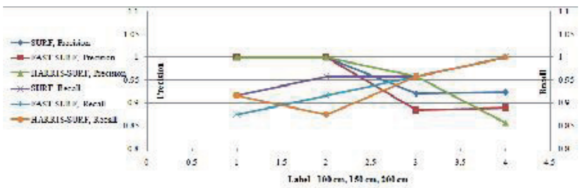

Figure 5. Precision and recall- 100 cm, 150 cm, 200 cm.

SURF method is helpful in detection and description steps as discussed in previous sections. The number of detected keypoints has been calculated which is shown for different distances in Table 10. Moreover, representation of keypoints is shown in Figure 6. Also Table 11, Table 12, and Table 13 represent needed time for implemented system.

Number of keypoints according to method and distance	SURF method
50 cm	9035
75 cm	4894
100 cm	7099

Table 10. Number of keypoints

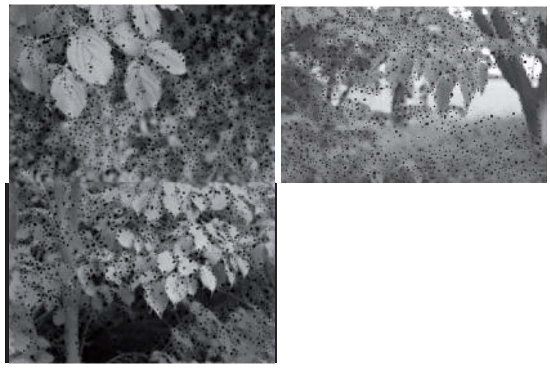

Figure 6. (Top left) Representation of keypoints for SURF method in distance 50 cm, (Top right) Representation of keypoints for SURF method in distance 75 cm, (Down left) Representation of keypoints for SURF method in distance 100 cm.

Distance	SURF Detector	SURF Descriptor	Operator SURF	Rest of Test Step
25 cm (80 Images)	301.683	582.2428	883.9258	0.009433
50 cm (80 Images)	317.5301	867.7071	1185.2372	0.008696
75 cm (80 Images)	328.998	1066.009	1395.007	0.0088608
100, 150, 200 cm (96 Images)	427.2995	1591.538	2018.8375	0.012067

Table 11. Needed time for the system used SURF

Distance	FAST Detector	SURF Descriptor	Operator FAST-SURF	Rest of Test Step
25 cm (80 Images)	0.483436	183.4367	183.920136	0.008029
50 cm (80 Images)	0.604643	269.0538	269.658443	0.008125
75 cm (80 Images)	0.677999	266.051	266.728999	0.007488
100, 150, 200 cm (96 Images)	1.059732	287.218	288.277732	0.011386

Table 12. Needed time for the system used FAST-SURF

Distance	HARRIS Detector	SURF Descriptor	Operator HARRIS-SURF	Rest of Test Step
25 cm (80 Images)	39.98336	182.3574	222.34076	0.007809
50 cm (80	40.11504	268.3261	308.44114	0.007712

Images)				
75 cm (80 Images)	40.11931	267.249	307.36831	0.006547
100, 150, 200 cm (96 Images)	48.08637	285.847	333.93337	0.009801

Table 13. Needed time for the system used HARRIS-SURF

When detection of keypoints has been done by using FAST method, the result is increase of computational speed. Although the combined FAST-SURF method has a good performance, it is robust in comparison to FAST-SIFT method when the distance is 25 cm. Its accuracy is also good enough to be applied in different systems. The needed time for the system with FAST-SURF is less than the system [15] which used SIFT as its component. The other used method is HARRIS-SURF. In this method, keypoints detection has been performed by using HARRIS method. In this case, number of detected keypoints is less than other two methods. The needed time is fair when this method has been applied. Another point is its accuracy. Its accuracy is the lowest one among all methods.

The system has the highest accuracy when it is implemented by SURF method. The SURF implemented system works well and has a good performance. As mentioned, SURF uses fast second order box filters to detect keypoints. In comparison to SIFT method in [15], SURF method detects more keypoints and the needed time has been increased. In addition to, SIFT uses smaller descriptors in comparison to SURF method which can lead to the mentioned fact. In detection step, SURF uses fast computation by means of box filters to have approximate Laplacian of Gaussian images if a fixed number of keypoints has been detected. If the number of keypoints is fixed for both SURF and SIFT algorithms, execution time is lower by using SURF, hence the needed time is comparable to the SIFT's needed time in [15]. There is also a good trade-off between time and accuracy in three implemented systems when they are compared to three other implemented systems in [15]. Additionally, proposed systems in [10] and [11] have been tested on artificial dataset while the used dataset in this work is complex and natural. Performance of the implemented systems is better than the proposed system in [9].

5. CONCLUSION

In this work, three different systems have been implemented by using combined methods of SURF, FAST, and HARRIS algorithms. In order to test the implemented systems, a very challenging dataset has been used. The dataset is composed of natural images with different point of views, angles, illumination, distances, and even weather conditions. Due to characteristics and properties of the dataset's images, the implemented systems have generality for plant recognition system. The results have been explained in details by conducting some experiments. Due to the performance of the implemented systems, they can be applied by considering the desired parameters. When the needed time should be lower, the FAST-SURF system can be applied. Obtained accuracy of the system is higher when SURF algorithm has been applied in the developed system. The next step of this work can be implementation of deep neural network to recognize plant species.

6. ACKNOWLEDGMENTS

The used dataset belongs to the Institute of Real-time Learning Systems, University of Siegen, Germany. It has been created and prepared by Masoud Fathi Kazerouni.

7. REFERENCES

[1] Du, J. X., Wang, X. F., & Zhang, G. J. (2007). Leaf shape based plant species recognition. Applied mathematics and computation, 185(2), 883-893.

[2] Ye, Y., Chen, C., Li, C. T., Fu, H., & Chi, Z. (2004, October). A computerized plant species recognition system. In Intelligent Multimedia, Video and Speech Processing, 2004. Proceedings of 2004 International Symposium on (pp. 723-726). IEEE.

[3] Zhenjiang, M., Gandelin, M. H., & Baozong, Y. (2006). An OOPR-based rose variety recognition system. Engineering Applications of Artificial Intelligence, 19(1), 79-101.

[4] Plotze, R. D. O., Falvo, M., Pádua, J. G., Bernacci, L. C., Vieira, M. L. C., Oliveira, G. C. X., & Bruno, O. M. (2005). Leaf shape analysis using the multiscale Minkowski fractal dimension, a new morphometric method: a study with Passiflora (Passifloraceae). Canadian Journal of Botany, 83(3), 287-301.

[5] Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y. X., Chang, Y. F., & Xiang, Q. L. (2007, December). 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). IEEE.

[6] Mallah, C., Cope, J., & Orwell, J. (2013). Plant leaf classification using probabilistic integration of shape, texture and margin features. Signal Processing, Pattern Recognition and Applications, 5, 1.

[7] Gouveia, F., Filipe, V., Reis, M., Couto, C., & Bulas-Cruz, J. (1997, July). Biometry: the characterisation of chestnut-tree leaves using computer vision. In Industrial Electronics, 1997. ISIE'97., Proceedings of the IEEE International Symposium on (pp. 757-760). IEEE.

[8] Saitoh, T., & Kaneko, T. (2000). Automatic recognition of wild flowers. In Pattern Recognition, 2000. Proceedings. 15th International Conference on (Vol. 2, pp. 507-510). IEEE.

[9] Mishra, P. K., Maurya, S. K., Singh, R. K., & Misra, A. K. (2012, March). A semi automatic plant identification based on digital leaf and flower images. In Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on (pp. 68-73). IEEE.

[10] Kazerouni, M. F., Schlemper, J., & Kuhnert, K. D. (2015). Comparison of modern description methods for the recognition of 32 plant species. Signal & Image Processing, 6(2), 1.

[11] Kazerouni, M. F., Schlemperz, J., & Kuhnert, K. D. (2015). Efficient Modern Description Methods by Using SURF Algorithm for Recognition of Plant Species. Advances in Image and Video Processing, 3(2), 10.

[12] Bay, H., Tuytelaars, T., & Van Gool, L. (2006). Surf: Speeded up robust features. Computer vision–ECCV 2006, 404-417.

[13] Harris, C., & Stephens, M. (1988, August). A combined corner and edge detector. In Alvey vision conference (Vol. 15, No. 50, pp. 10-5244).

[14] Rosten, E. & Drummond, T. (May 2006). Machine Learning for High-Speed Corner Detection. In European Conference on Computer Vision (vol. 1, pp. 430–443).

[15] Kazerouni, M. F., Schlemper, J., & Kuhnert, K. D. Modern Detection and Description Methods for Natural Plants Recognition (Periodical style—Accepted), ICMLA 2017: 19th International Conference on Machine Learning and Applications, Copenhagen, Denmark.

[16] https://en.wikipedia.org/wiki/Speeded_up_robust_features

[17] Hossain, J., & Amin, M. A. (2010, December). Leaf shape identification based plant biometrics. In Computer and Information Technology (ICCIT), 2010 13th International Conference on (pp. 458-463). IEEE.

[18] Vapnik, V. (1995). The natural of statistical theory.

[19] Osuna, E., Freund, R., & Girosi, F. (1997, September). An improved training algorithm for support vector machines. In Neural Networks for Signal Processing [1997] VII. Proceedings of the 1997 IEEE Workshop (pp. 276-285). IEEE.

[20] Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. Advances in neural information processing systems, 9, 155-161.

[21] Priya, C. A., Balasaravanan, T., & Thanamani, A. S. (2012, March). An efficient leaf recognition algorithm for plant classification using support vector machine. In Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on (pp. 428-432). IEEE.

[22] https://en.wikipedia.org/wiki/Grayscale#Converting_color_to_grayscale

[23] Bianco, S., Mazzini, D., Pau, D. P., & Schettini, R. (2015). Local detectors and compact descriptors for visual search: a quantitative comparison. Digital Signal Processing, 44, 1-13.

[24] Jégou, H., Perronnin, F., Douze, M., Sánchez, J., Perez, P., & Schmid, C. (2012). Aggregating local image descriptors into compact codes. IEEE transactions on pattern analysis and machine intelligence, 34(9), 1704-1716.

[25] Moravec, H. P. (1977). Towards automatic visual bbstacle avoidance. In International Conference on Artificial Intelligence (5th: 1977: Massachusetts Institute of Technology).

[26] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 1, pp. I-I). IEEE.

[27] Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). Computer vision and image understanding, 110(3), 346-359.

[28] Brown, M., & Lowe, D. G. (2002, September). Invariant Features from Interest Point Groups. In BMVC (Vol. 4).

[29] Caicedo, J. C., Cruz, A., & Gonzalez, F. A. (2009, July). Histopathology image classification using bag of features and kernel functions. In Conference on Artificial Intelligence in Medicine in Europe (pp. 126-135). Springer Berlin Heidelberg.

[30] Jégou, H., Douze, M., & Schmid, C. (2010). Improving bag-of-features for large scale image search. International journal of computer vision, 87(3), 316-336.

[31] Rahman, M. M., Antani, S. K., & Thoma, G. R. (2011, June). Biomedical CBIR using "bag of keypoints" in a modified inverted index. In Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on (pp. 1-6). IEEE.

[32] MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability (Vol. 1, No. 14, pp. 281-297).

[33] Purohit, S., Viroja, R., Gandhi, S., & Chaudhary, N. (2015, December). Automatic plant species recognition technique using machine learning approaches. In Computing and Network Communications (CoCoNet), 2015 International Conference on (pp. 710-719). IEEE.

[34] http://scikit-learn.org/stable/modules/model_evaluation.html#accuracy-score