

RESEARCH ARTICLE

# A Novel Method of Automatic Plant Species Identification Using Sparse Representation of Leaf Tooth Features

Taisong Jin<sup>1</sup>, Xueliang Hou<sup>2\*</sup>, Pifan Li<sup>1</sup>, Feifei Zhou<sup>1</sup>

**1** School of Information Science and Engineering, Xiamen University, Xiamen, 361005, China, **2** School of Life Sciences, Xiamen University, Xiamen, 361005, China

\* [houxueliang@xmu.edu.cn](mailto:houxueliang@xmu.edu.cn)



## OPEN ACCESS

**Citation:** Jin T, Hou X, Li P, Zhou F (2015) A Novel Method of Automatic Plant Species Identification Using Sparse Representation of Leaf Tooth Features. PLoS ONE 10(10): e0139482. doi:10.1371/journal.pone.0139482

**Editor:** Zhong-Jian Liu, The National Orchid Conservation Center of China; The Orchid Conservation & Research Center of Shenzhen, CHINA

**Received:** January 12, 2015

**Accepted:** September 13, 2015

**Published:** October 6, 2015

**Copyright:** © 2015 Jin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

Automatic species identification has many advantages over traditional species identification. Currently, most plant automatic identification methods focus on the features of leaf shape, venation and texture, which are promising for the identification of some plant species. However, leaf tooth, a feature commonly used in traditional species identification, is ignored. In this paper, a novel automatic species identification method using sparse representation of leaf tooth features is proposed. In this method, image corners are detected first, and the abnormal image corner is removed by the PauTa criteria. Next, the top and bottom leaf tooth edges are discriminated to effectively correspond to the extracted image corners; then, four leaf tooth features (Leaf-num, Leaf-rate, Leaf-sharpness and Leaf-obliqueness) are extracted and concatenated into a feature vector. Finally, a sparse representation-based classifier is used to identify a plant species sample. Tests on a real-world leaf image dataset show that our proposed method is feasible for species identification.

## Introduction

Because of a shortage of skilled subject matter experts, traditional species identification is tedious and difficult. The development of ubiquitous technologies, such as digital cameras and portable computers, has made automatic species identification possible. Specifically, for a large scale dataset of plant species, traditional species identification requires significant manual effort. It is unlikely that traditional species identification methods are practical; therefore, automatic species identification methods are more and more crucial. Leaf characters, involving shape, are used extensively in traditional text-based taxonomic keys for plant identification, which leads leaf shape to be the first and most commonly studied leaf feature.

## Leaf shape feature

Ingrouille's work [1], for example, is one of the earliest efforts in automatic species identification, where twenty-seven leaf shape features were extracted to identify oak species. The following five

shape features are also used for species identification: Fourier descriptors [2,3], contour signatures [4,5], landmark [6,7], quantitative shape descriptors [8,9] and fractal dimensions [10].

### Leaf venation feature

The second most commonly studied feature is leaf venation, also referred to as the leaf Vein. Leaf venation provides leaves with a specific structure, which may be useful to identify some species. The key is to effectively extract leaf venation; in this regard, several methods [11,12] have been proposed to effectively extract leaf venation. However, there have been fewer attempts to analyze or compare leaf venation; furthermore, most methods exploited synthetic or manually extracted leaf venation images.

### Leaf texture feature

The third most commonly studied feature is leaf texture; different leaf texture extraction methods are proposed in [13,14]. Although the existing leaf texture based methods have shown promising performance for some species, experiments using those methods involved very limited datasets. Thus, it is difficult to evaluate and is impossible to scale up a large scale dataset.

### Leaf margin feature

The fourth most commonly studied feature is leaf margin. Despite being a useful feature of leaves in traditional species identification, leaf margin has seen very little use in automatic species identification due to the difficulty in acquiring quantitative measurements automatically. Currently, only a few methods using leaf margin are proposed. Clark et al. [15], for example, used manually taken measurements, such as tooth length and width, to help automate species identification. Zheng et al. [16] extracted three morphological leaf tooth measurements, namely the number, sharpness and inclination, for plant identification. Zheng et al. [17] also defined a function to extract leaf lobe features. Cope's study [18] gives a recent comprehensive review on automatic species identification.

All the above methods demonstrate that automatic species identification is suitable for some species. However, the key issue to solve the problem of robust automatic species identification is how to deal with the different deformation of leaf character and the large and small inter-class variations that are typical of botanical samples. Even if the study focuses on a single genus, it may contain many species, each of which encompasses extensive variation between constituent populations. Thus, the existing methods are insufficient to identify the complex species; additional features must be incorporated into the current automatic species identification method.

On the other hand, the classifier is very crucial for obtaining promising performance of species identification. Existing classifiers, such as *K* Nearest Neighbor (*K*-NN) [9], Random Forest [11], Support Vector Machine (SVM) [19] have been employed to identify plant species. More recently, the sparse representation based classifier has shown promising performance in face recognition [20], image analysis [21], and other applications [22,23]. However, to the best of our knowledge, the classifier based on sparse representation has not yet been applied to plant species identification.

Inspired by the recent progress of species identification and the sparse representation based classifier, we propose, in this article, a novel automatic plant species identification method. Unlike existing methods, our proposed method is based exclusively on leaf tooth; whereas, leaf shape, venation and texture are discarded. The contributions of this paper are as follows:

1. The morphological measurements of four leaf tooth features are proposed. Our proposed measurements effectively distinguish between the top and bottom edges of a leaf tooth. In addition, the noise effect is also removed using the PauTa criteria. Thus, our method is more suitable for real-world applications.
2. A sparse representation based classifier is applied to plant species identification. In our proposed method, an overall dictionary is constructed, and the species of a test sample is decided by the projection coefficients in the dictionary.
3. To demonstrate the feasibility of our proposed method, we conducted experiments on a real-world plant species dataset. In particular, we compared our proposed method with the  $K$  nearest neighbor ( $K$ -NN)-based and BP neural network-based methods.

## Materials and Methods

### Image pre-processing

A digital image of a plant leaf is usually obtained by a digital camera or a scanner. Compared with a scanner, digitalization using a digital camera is more suitable for image processing. Thus, in our experiments, we used a digital camera to digitalize plant leaf images. Because leaves are rarely perfectly flat and are affected by shadows and noise, we applied, to the leaf image, pre-processing carried out as follows: Firstly, the color image was converted into a gray-scale image. Secondly, a binary image was obtained via adaptive image thresholding on the obtained grayscale image. A Roberts cross operator was further applied to the binary image to obtain an edge image. Thirdly, in order to remove many non-leaf margin edges retained in the edge image (specifically in the region of leaf venation) the dilation operator in morphology operations was used to fill in these holes. In our experiments, the dilation operator effectively removed most non-leaf-margin edges. Finally, the thin operator in mathematical morphology operations was used to make the leaf edge as thin as one pixel.

All the aforementioned operations were implemented using the Matlab Toolbox. As shown in Figs 1, 2, 3, our proposed image pre-processing operations effectively extracted leaf margins.

### Extraction of image corners

Noise and illumination do not easily obscure image corners. Thus, image corners are commonly used to correspond to leaf teeth in the image. In our method, we also apply image corners to correspond to leaf teeth. For the different image corner extraction methods, researchers [16,24] have shown that Susan corner detection is more resistant to the effect of noise and is usually superior to other methods. Thus, we applied Susan corner detection to extract image corners from edge images. After extracting the image corners, most image corners corresponding to leaf teeth were detected. Due to the effect of noise, a few abnormal image corners may not correspond to leaf teeth. Thus, these abnormal image corners must be removed.

### The removal of abnormal image corners

We exploited the PauTa criteria to remove non-leaf-tooth image corners. For convenience, important notations are listed in Table 1. For each image corner,  $C_i$ , we calculated two formulas according to the following PauTa criteria:  $\|C_i - \mu_1\| > 3\sigma_1$  and  $\|C_i - \mu_2\| > 3\sigma_2$ , where  $\mu_1$  and  $\mu_2$  are the means of horizontal coordinates and vertical coordinates of the image corners, respectively.

To avoid one, or several, measurements of larger variation, abnormal image corners were removed based on the aforementioned Pauta criterion, i.e, if  $\|C_i - \mu_1\| > 3\sigma_1$  or  $\|C_i - \mu_2\| > 3\sigma_2$ , then  $C_i$  was considered an abnormal image corner and was removed; otherwise, if  $\|C_i - \mu_1\| <$



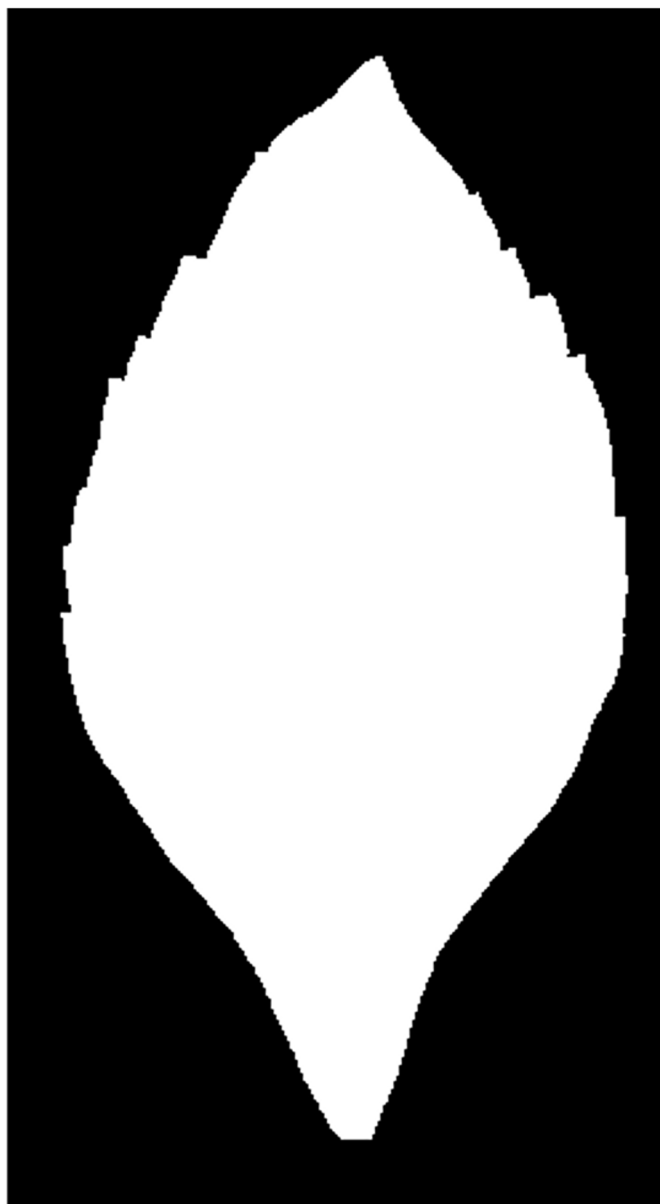
**Fig 1. An example image for *Duranta repens* Linn.**

doi:10.1371/journal.pone.0139482.g001

$3\sigma_1$  and  $||C_i - \mu_2|| < 3\sigma_2$ , then  $C_i$  was considered a normal image corner and was preserved. Generally, zero to two abnormal image corners were removed in the implementation. Note that the operators 'and' and 'or' in the formulas are logical operators.

### Distinguishing the top and bottom edges of leaf teeth

After detection, image corners further corresponded to the bottom or top edges of leaf teeth. Firstly, a simple polygon was obtained by connecting the image corners with straight lines. Secondly, these image corners were distinguished in terms of the inner angle of a simple polygon,



**Fig 2. The obtained binary image.**

doi:10.1371/journal.pone.0139482.g002

i.e., if the inner angle of the polygon is smaller than  $180^\circ$ , then the image corner vertex corresponds to the top edge of the leaf tooth; otherwise, it corresponds to the bottom edge. As shown in Figs 4 and 5, our proposed method correctly extracts the leaf teeth and also discriminates the top and bottom edges of the leaf teeth.

### Measurements of Leaf Tooth Features

After the leaf teeth are detected, the morphological measurements of four leaf tooth features are presented. In this section, we adopt the following four measurements as leaf tooth features:

**The first feature**, Leaf-num, is denoted by the first measure according to the total number of leaf teeth described by [16];



**Fig 3. The obtained edge image via edge operator and thin operation.**

doi:10.1371/journal.pone.0139482.g003

**The second feature**, Leaf-rate, is denoted by the second measure according to the ratio between the number of leaf teeth and the length of the leaf margin expressed in pixels.

**The third feature**, Leaf-sharpness, is measured as follows: For each leaf tooth, an acute triangle is obtained by connecting the top edge and two bottom edges of the leaf tooth. Thus, for a leaf image, many triangles corresponding to leaf teeth are obtained. In our method, the acute angle for each leaf tooth is exploited as a measure for plant identification. In detail, the third

**Table 1. List of important notations.**

	Notation and Description
$\mu_1$	The mean of horizontal coordinates of the image corners
$\sigma_1$	The standard deviation of horizontal coordinates of the image corners
$\mu_2$	The mean of vertical coordinates of the image corners
$\sigma_2$	The standard deviation of vertical coordinates of the image corners
$\mu_3$	The mean of the third measures of all the leaf teeth
$\sigma_3$	The standard deviation for the third measures of all the leaf teeth
$\mu_4$	The mean of the fourth measures of all the leaf teeth
$\sigma_4$	The standard deviation for the third measures of all the leaf teeth
$Ta_i$	The third measure of the i-th leaf tooth
$Te_i$	The fourth measure of the i-th leaf tooth
$D_i$	The i-th species-specific sub-dictionary
$D$	The overall dictionary
$Y$	The sample to be identified
$X_i$	The projection coefficient on the i-th species-specific sub-dictionary
$X$	The projection coefficient on the overall dictionary

doi:10.1371/journal.pone.0139482.t001

measure of the i-th leaf tooth, namely,  $Ta_i$ , is calculated by the following formula:

$$Ta_i = \arccos\langle \mathbf{a}, \mathbf{b} \rangle, \quad (1)$$

where  $\mathbf{a}, \mathbf{b}$  are the vectors obtained by connecting the top edge and two bottom edges of the i-th leaf tooth, respectively. In order to resist the effect of noise, we firstly remove the abnormal measure of the leaf teeth in terms of PauTa criteria as follows: the mean,  $\mu_3$ , and standard deviation,  $\sigma_3$ , of all the third measures of the leaf teeth are calculated; next, for the i-th acute angle,  $Ta_i$  ( $i \in \{1, 2, \dots, m\}$ ), if  $||Ta_i - \mu_3|| > 3\sigma_3$ , then  $Ta_i$  is considered to be an abnormal measure and is removed. Otherwise,  $Ta_i$  is preserved for further processing. All the preserved measures are used for re-calculating the new mean as the third leaf tooth feature.

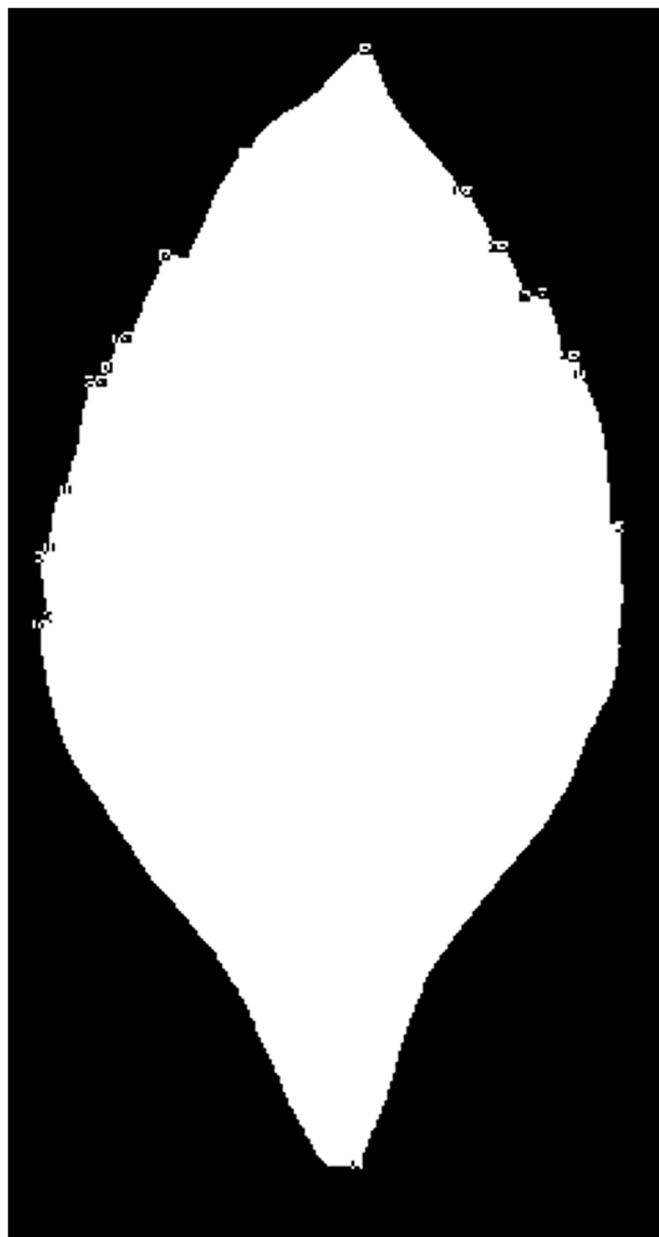
**The fourth feature**, Leaf-obliqueness, is measured as follows: Firstly, the fourth measure for the i-th leaf tooth,  $Te_i$ , is formulated as

$$Te_i = \frac{|d_1|}{|d_2|} = \frac{|d_1|^2}{2 \times S}, \quad (2)$$

where  $|d_1|$  and  $|d_2|$  are the acute triangle's base and height, respectively, obtained by the i-th leaf tooth;  $S$  is the area of the triangle. As such, the effect of noise leads to abnormal measures. Thus, by the PauTa criteria, the abnormal measures must be removed. The details of how to remove the abnormal measures are as follows: Firstly, the mean,  $\mu_4$ , and the standard deviation,  $\sigma_4$ , of all the fourth measures of the leaf teeth are calculated; Secondly, for the i-th acute angle,  $Te_i$ . ( $i \in \{1, 2, \dots, m\}$ ), if  $||Te_i - \mu_4|| > 3\sigma_4$ , then  $Te_i$  is considered as an abnormal angle and removed; otherwise,  $Te_i$  is preserved for further processing. All the preserved measures are used to re-calculate the new mean as the fourth leaf tooth feature. The third and fourth features for one leaf tooth are illustrated in Fig 6.

## The Proposed Sparse Representation Based Classifier

After the four leaf tooth features are computed, they are concatenated into a feature vector. Then, a classifier is used to classify the feature vector as one plant species. Recently, a sparse representation-based classifier has shown promising performance for many applications. In



**Fig 4. The image corners detected by the Susan method for *Duranta repens* Linn.**

doi:10.1371/journal.pone.0139482.g004

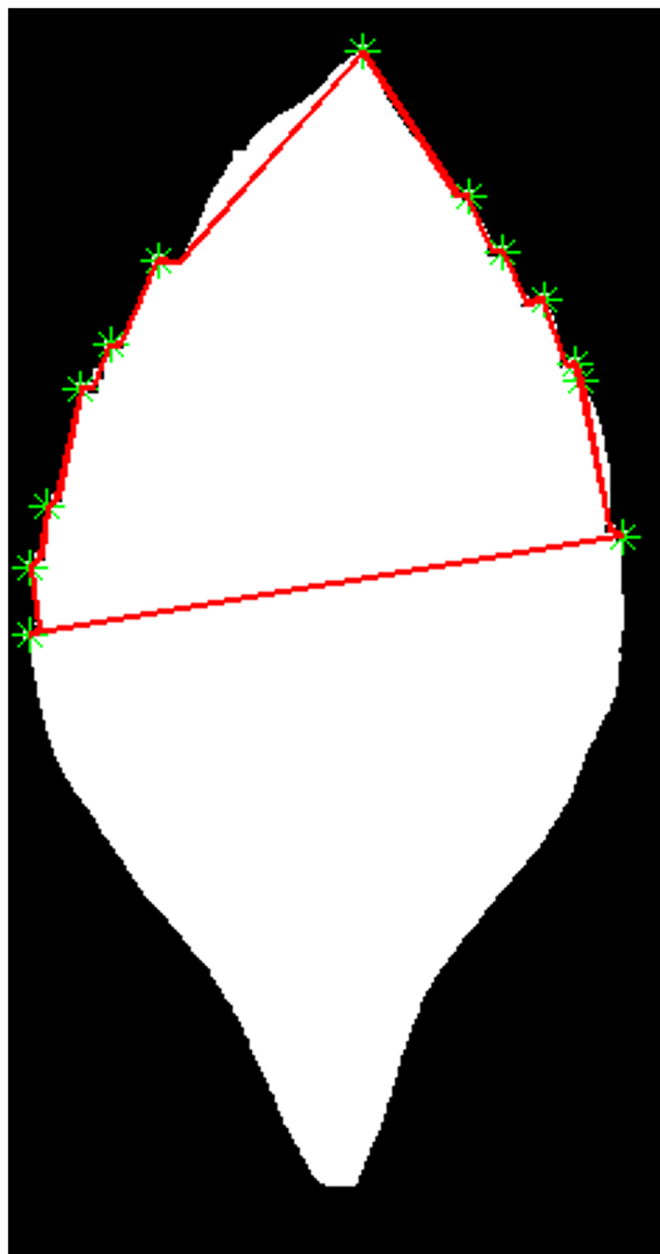
our method, we incorporate a sparse representation-based classifier into plant species identification.

### Plant identification using the sparse representation based classifier

Sparse representation represents a signal as a linear combination of as few atoms as possible of a given dictionary. In other words, given the signals  $\mathbf{y} \in \mathbb{R}^n$  and  $\mathbf{D} \in \mathbb{R}^{n \times m}$ , the sparse representation problem is formulated as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0, \text{ s. t. } \mathbf{y} = \mathbf{D}\mathbf{x}, \quad (3)$$





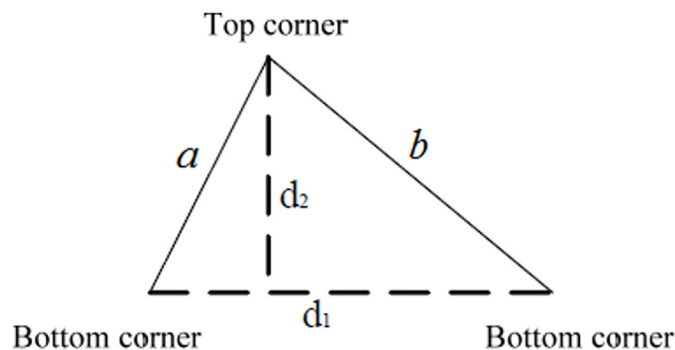
**Fig 5. The image corners responding to leaf teeth.**

doi:10.1371/journal.pone.0139482.g005

where  $\|\mathbf{X}\|_0$  is the  $\ell_0$  norm of the coefficient vector,  $\mathbf{x} \in \mathbb{R}^m$ , i.e., the number of non-zero elements. The optimization of Eq (3) is a NP-hard problem, and direct optimization is not feasible. This problem is usually solved approximately by replacing  $\ell_0$  norm with  $\ell_1$  norm. Then, Eq (3) is transformed into the following optimization problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1, \text{ s. t. } \mathbf{y} = \mathbf{D}\mathbf{x}. \quad (4)$$

The key issue of sparse representation is to obtain an over-completed dictionary that allows the sparse representation of the data samples with respect to the obtained dictionary. Many methods have been proposed to obtain such a dictionary. For instance, in [20], the entire



**Fig 6. Illustration of the third and fourth leaf tooth features.**

doi:10.1371/journal.pone.0139482.g006

training set is used as a dictionary, where excellent performance is shown for some real-world face image datasets. However, such a dictionary is not practical for large-scale image datasets. In [21], the visual feature vectors extracted from the images are used to build a dictionary, which significantly decreases the computational load. In the field of species identification, because of the large computational cost, it is unsuitable to use the whole training image set to obtain a dictionary. Thus, to build the dictionary, our proposed method applies the leaf tooth features extracted from the training set.

Suppose the training dataset consists of leaf images of  $q$  plant species, with  $p_i$  samples from the  $i$ th species; the total number of samples is  $n = \sum_{i=1}^q p_i$ . Firstly, the four leaf tooth features are

calculated for each image. Furthermore, the four leaf tooth features are concatenated into a feature vector. Thus, a set of feature vectors is obtained for the entire training dataset. Let the feature vector extracted from the  $n$ -th image of the  $m$ -th plant species be  $\mathbf{d}_{mn}$ . Specifically, let all the feature vectors corresponding to  $m$ -th plant species be concatenated into a matrix  $\mathbf{D}_m = [\mathbf{d}_{m1} \mathbf{d}_{m2} \cdots \mathbf{d}_{mp_m}]$ , namely, the  $m$ -th species-specific sub-dictionary. Thus, the overall dictionary  $\mathbf{D} = [\mathbf{D}_1 \mathbf{D}_2 \cdots \mathbf{D}_q]$  is built by concatenating all the species-specific sub-dictionaries.

Given an overall dictionary, sparse representation means that any sample is represented linearly by a few atoms of the dictionary according to Eq (3), and a projection coefficient vector is obtained.

Ideally, for the projection coefficient vector  $\mathbf{x} = [000 \cdots x_{i1} x_{i2} \cdots x_{im} \cdots 000]$  of the sample on the dictionary, most of the coefficients are zero; only as few coefficients as possible, corresponding to its species-specific sub-dictionary, are not zero. In practice, because of the effect of noise, some coefficients, corresponding to other species-specific sub-dictionaries, are also non-zero. The re-constructed residuals on different species-specific sub-dictionaries are usually different. The smaller the residual on the species-specific sub-dictionary, the more likely it is that the sample belongs to the corresponding species. Thus, plant species can be identified according to the residuals on species-specific sub-dictionaries.

To remove the effect of noise, projection coefficients are computed on the dictionary according to Eq (4), i.e., species identification is transformed into the following optimization problem:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s. t. } \|\mathbf{y} - \mathbf{D}\mathbf{x}\|^2 \leq \varepsilon, \quad (5)$$

where  $\varepsilon$  is the threshold of the residual. This problem is transformed into the following problem:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|^2 + \alpha \|\mathbf{x}\|_1, \quad (6)$$

where the parameter,  $\alpha$ , indicates the trade-off between the sparsity of the solution and the reconstruction error. This is a constrained lasso problem; the detailed solution of Eq (6) is found in [23].

Given a test sample,  $y$  (to be identified), the method in [23] for solving the Lasso problem in Eq (6) is performed to obtain the projection coefficients. Additionally, the residuals on the species-specific sub-dictionaries are formulated as

$$r^m(y) = \|y - D_m x_m\|^2, \quad m = 1, 2, \dots, q, \quad (7)$$

where  $x_i$  is the projection coefficient on the  $i$ -th species-specific sub-dictionary. Then, the species of the test sample,  $y$ , is identified as

$$\text{Species}(y) = \min_{m=1,2,\dots,q} r^m(y). \quad (8)$$

The proposed plant identification method has the following advantage: For adding a new plant species, our proposed method needs to only incorporate the plant samples of the corresponding species into an overall dictionary. Thus, it is easy to extend more plant species to realize the species identification task.

Our proposed method is described as follows:

Input: a normalized overall dictionary,  $D'$ , and a test sample,  $y$ .

Output: the species of the test sample.

Step 1: Normalize the test sample.

Step 2: Calculate the projection coefficients of the test sample on an overall dictionary according to Eq (6).

Step 3: Calculate the re-constructed residuals on species-specific dictionaries according to Eq (7).

Step 4: Output the species of the test sample according to Eq (8).

## Implementation

To effectively implement our proposed automatic identification procedure, all the morphological measurements of the four leaf tooth features are first normalized to the range [0,1]. Then, to normalize an overall dictionary, the morphological measurements are scaled to be unitary. Furthermore, the corresponding features of the test image are also normalized to the range [0,1].

## Results, Discussion, and Conclusions

### The experimental setting

Because different species own leaf characteristics with large and small inter-class variations, even if the study focuses on a single genus, it may contain many species, each of which encompasses much variation between constituent populations. Thus, we specifically select the species, which have similar leaf tooth character and observe the classification performance of the proposed method. To demonstrate the performance of our proposed method, we constructed, as an image dataset, a total of 700 leaf images (S1 File) from eight plant species. For each species, there are leaf images with variations in lighting, scale and background. The eight species have images as follows: (1) 54 images of *Hibiscus rosa-sinensis* Linn; (2) 96 images of *Duranta repens* Linn; (3) 54 images of *Parthenocissus tricuspidata* (Sieb. et Zucc.) Planch; (4) 124 images of *Hibiscus schizopetalus* (Masters) Hook. f; (5) 100 images of *Cyclobalanopsis glauca* (Thunb.) Oerst; (6) 82



**Fig 7. The example image for *Hibiscus rosa-sinensis* Linn.**

doi:10.1371/journal.pone.0139482.g007

images of *Eriobotrya japonica* (Thunb.) Lindl; (7) 124 images of *Conyza canadensis* (L.) Cronq; and (8) 66 images of *Amygdalus persica* Linn. Example images are shown in Figs [7](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#). A total of 350 images were used as the training dataset, where half of the images were randomly selected for each species. Then, the other 350 images were used as the test dataset.

### The plant identification test

Using our method, we first extracted the leaf tooth features from the training dataset and further built the overall dictionary; then, we ran our proposed method on the test dataset. The



**Fig 8. The example image for *Duranta repens* Linn.**

doi:10.1371/journal.pone.0139482.g008



**Fig 9. The example image for *Parthenocissus tricuspidata* (Sieb. et Zucc.) Planch.**

doi:10.1371/journal.pone.0139482.g009

experimental results are listed in [Table 2](#). The identification accuracy per species is reported over five independent runs, where the accuracy per species is the percentage of the successful identifications of each species relative to the number of leaves from the corresponding species.

According to the experimental results (See [Table 2](#)), our proposed method achieved the following classification accuracies in specie identification: (1) *Hibiscus rosa-sinensis* Linn: 75.0%; (2) *Duranta repens* Linn: 79.3%; (3) *Parthenocissus tricuspidata* (Sieb. et Zucc.) Planch: 76.3%; (4) *Hibiscus schizopetalus* (Masters) Hook. f.: 76.6%; (5) *Cyclobalanopsis glauca* (Thunb.) Oerst.: 77.3%; (6) *Eriobotrya japonica* (Thunb.) Lindl.: 75.5%; (7) *Conyza canadensis* (L.) Cronq.: 74.7%; and (8) *Acalypha wilkesiana* Muell.-Arg.: 72.8%. The experimental results in [Table 2](#) show that our proposed method achieves an average classification rate greater than 76% for all species.

The classification performance should be significantly greater. A possible explanation is that the similar characteristics of leaf teeth of different species, to some extent, affect the classification performance of our proposed method. However, our proposed method is based exclusively on leaf teeth; leaf shape, veining and texture were discarded. We consider that the identification results are acceptable for automatic species identification. Moreover, the classification results can be further enhanced by introducing other leaf features such as leaf shape and leaf venation. Because we focused on leaf tooth features in this paper, we didn't conduct tests using other leaf features.

### The statistics of leaf tooth features

The four leaf tooth features extracted from the example images (See [Fig 4](#)) are shown in [Table 3](#). From the data in [Table 3](#), we conclude the following: the leaf tooth features obtained by our





**Fig 10.** The example image for *Hibiscus schizopetalus* (Masters) Hook. f.

doi:10.1371/journal.pone.0139482.g010



**Fig 11.** The example image for *Cyclobalanopsis glauca* (Thunb.) Oerst.

doi:10.1371/journal.pone.0139482.g011



**Fig 12.** The example image for *Eriobotrya japonica* (Thunb.) Lindl.

doi:10.1371/journal.pone.0139482.g012

proposed method are intuitively consistent with human observations. Thus, because traditional plant identification commonly considers leaf tooth to be effective, the four leaf tooth features obtained by our proposed method are feasible. It is to be emphasized here that the results obtained by our proposed method are not consistent with the real results; however, the identification test we conducted demonstrated that our method is effective for plant identification.

### The comparison with the other methods

To further compare our proposed method with existing methods, we applied two existing classifiers to the same training and test datasets. The common classifiers are the  $K$ -NN and BP Neural Network used for plant identification applications.  $K$ -NN is a non-parametric classifier, and the test sample is identified by a majority vote of its neighbors, with the sample being assigned to the species that is most common among its  $K$  nearest neighbors. BP neural network is a multiple-layer feed-forward network trained according to the error back propagation algorithm. In our test, the  $K$  Nearest Neighbor parameter is set to 2, and the BP Neural Network consists of four inputs, one output, and one hidden layer composed of ten neuron units; transfer functions of each unit adopt a logarithmic sigmoid transfer function.

The classification accuracy per species for all three classifiers, where the accuracy per species is the percentage of the successful identifications of each species relative to the number of leaves





**Fig 13.** The example image for *Conyza canadensis* (L.) Cronq.

doi:10.1371/journal.pone.0139482.g013

from the corresponding species, are reported over five independent runs. The species identification results for the classifiers are listed in [Table 4](#).

From the experimental results shown in [Table 4](#), we observe the following: our proposed method outperforms the  $K$ -NN and BP neural network methods. By comparison, the total accuracy of species identification of our proposed method is improved by about 4% and 3%, respectively. The possible reason is that the sparse representation based classifier is more robust to the effect of noise and other factors such as illumination and shadow. Thus, our proposed classifier is feasible for plant identification.



**Fig 14. The example image for *Acalypha wilkesiana* Muell.-Arg.**

doi:10.1371/journal.pone.0139482.g014

**Table 2. Accuracy of species identification of our proposed method.**

Species	Mean accuracy $\pm$ Standard deviation (%)
<i>Hibiscus rosa-sinensis</i> Linn	75.0 $\pm$ 3.4
<i>Duranta repens</i> Linn	79.3 $\pm$ 2.1
<i>Parthenocissus tricuspidata</i> (Sieb. et Zucc.) Planch	76.3 $\pm$ 3.2
<i>Hibiscus schizopetalus</i> (Masters) Hook. f.	76.6 $\pm$ 2.9
<i>Cyclobalanopsis glauca</i> (Thunb.) Oerst	77.3 $\pm$ 2.8
<i>Eriobotrya japonica</i> (Thunb.) Lindl.	75.5 $\pm$ 4.5
<i>Conyza canadensis</i> (L.) Cronq.	74.7 $\pm$ 1.7
<i>Acalypha wilkesiana</i> Muell.-Arg.	72.8 $\pm$ 3.6

doi:10.1371/journal.pone.0139482.t002

**Table 3.** The four leaf tooth features for four example images.

Species	leaf-num	leaf-rate	leaf-sharpness	leaf-obliqueness
<i>Hibiscus rosa-sinensis</i> Linn	26	0.0137	112.2600	0.2094
<i>Duranta repens</i> Linn	11	0.0068	105.5330	0.1438
<i>Parthenocissus tricuspidata</i> (Sieb. et Zucc.) Planch	12	0.0072	97.4062	0.2456
<i>Hibiscus schizopetalus</i> (Masters) Hook. f.	14	0.0086	101.4717	0.1242
<i>Cyclobalanopsis glauca</i> (Thunb.) Oerst	12	0.0152	140.1123	0.1169
<i>Eriobotrya japonica</i> (Thunb.) Lindl.	15	0.0138	133.2329	0.1342
<i>Conyza canadensis</i> (L.) Cronq.	23	0.0208	80.8092	0.1914
<i>Acalypha wilkesiana</i> Muell.-Arg.	80	0.0149	119.3086	0.2651

doi:10.1371/journal.pone.0139482.t003

**Table 4.** Species identification results using the different classifiers.

Classifiers	Mean accuracy ± Standard deviation (%)
BP neural network	73.2±3.3
K-NN	72.3±2.5
Our proposed classifier	76.3±3.4

doi:10.1371/journal.pone.0139482.t004

## Conclusion

In this paper, we proposed a novel automatic plant species identification method using sparse representation of leaf tooth features. Our proposed method introduced four leaf tooth features for plant species identification. Then, a sparse representation-based classifier was used to identify the plant species for a test sample. We conducted the experiments on a real-world dataset, showing that our proposed method was feasible for plant identification. In future work, we plan to study the more complex features of leaf teeth. In addition, other features such as shape, color and texture will be incorporated into the sparse representation based plant species classification framework. Thus, the classification performance can be significantly enhanced.

## Supporting Information

**S1 File. Plant Species dataset.**  
(RAR)

## Acknowledgments

The authors thank Michael McAllister for his assistance in proofreading the manuscript.

## Author Contributions

Conceived and designed the experiments: XLH TSJ. Performed the experiments: TSJ. Analyzed the data: PFL. Contributed reagents/materials/analysis tools: FFZ. Wrote the paper: FFZ.

## References

1. Ingrouille MJ, Laird SM. A quantitative approach to oak variability in some north London woodlands. *LON Nat*. 1986; 65:35–46.
2. Neto J, Meyer G, Jones D, Samal A. Plant species identification using elliptic Fourier leaf shape analysis. *Comput Electron Agr*. 2006; 50:121–134.

3. Lexer C, Joseph J, Loo MV, Prenner G, Heinze B, Chase MW, et al. The use of digital image-based morphometrics to study the phenotypic mosaic in taxa with porous genomes. *Taxon*.2009; 58:349–364.
4. Meade C, Parnell J., Multivariate analysis of leaf shape patterns in Asian species of the *Uvaria* group (Annonaceae). *Bot J Linn Soc*.2003; 143:231–242.
5. Mokhtarian F, Abbasi S. Matching shapes with self-intersection: application to leaf classification, *IEEE T Image Processing*.2004; 13:653–661.
6. Jensen RJ, Ciofani KM, Miramontes LC. Lines, outlines, and landmarks: morphometric analyses of leaves of *Acer rubrum*, *Acer saccharinum* (Aceraceae) and their hybrid. *Taxon*.2002; 51:475–492.
7. Corney DPA, Clark JYC, Tang HT, Wilkin P. Automatic extraction of leaf characters from herbarium specimens. *Taxon*.2012; 61:31–244.
8. Du JX, Wang XF, Zhang GJ. Leaf shape based plant species recognition. *Appl Math Comput*. 2007; 185(2):883–893.
9. Agarwal G., Belhumeur P., Feiner S., Jacobs D., Kress W. J., Ramamoorthi R., et al. First steps toward an electronic field guide for plants. *Taxon*.2006; 55, 597.
10. Du J, Zhai C, Wang Q. Recognition of plant leaf image based on fractal dimension features. *Neurocomputing*.2013; 116(20):150–156.
11. Larese MG, Namías R, Craviotto RM, Arango MR. Automatic classification of legumes using leaf vein image features. *Pattern Recogn*.2014, 47(1):158–168.
12. Park J, Hwang E, Nam Y. Utilizing venation features for efficient leaf image retrieval. *J Syst Software*.2008; 81 (1):71–82.
13. Backes AR, Gonçalves WN, Martinez AS, Bruno OM. Texture analysis and classification using deterministic tourist walk. *Pattern Recogn*,2010; 43, 685–694.
14. Ramos E, Fernandez DS. Classification of leaf epidermis microphotographs using texture features. *Ecol Inform*.2009; 4:177–181.
15. Clark JY. Neural networks and cluster analysis for unsupervised classification of cultivated species of *Tilia* (Malvaceae). *Bot J Linn Soc*.2009; 159:300–314.
16. Zheng X, Wang X, Gao J. Application of SUSAN algorithm in feature extraction of leaf margin. *Chin. Agric. Sci. Bull*.2011; 27:174–178.
17. Zheng X, Zhang X, Bo S. The study on automatic feature extraction of leaf lobes. *Chin. Agric. Sci. Bull*.2012; 28(27):152–156.
18. Cope JS, Corney D, Clark JY, Remagnino P, Wilkin P. Plant species identification using digital morphometrics: a review. *Expert Syst Appl*. 2012; 39:7562–7573.
19. Martins J,Oliveira LS, Nisgoski S, Sabourin R. A database for automatic classification of forest species. *MACH VISION APPL*; 2013, 24:567–578.
20. Wright J, Yang AY, Ganesh A, et al. Robust face-recognition via sparse representation. *IEEE T Pattern Anal Mach Intel*.2008; 31(2): 210–227.
21. Bao B, Zhu G, Shen J, Yan S. Robust image analysis with sparse representation on quantized visual features. *IEEE T Image Process*.2013; 22(3):860–871.
22. Bruckstein, Donoho D, Elad M.From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*.2009; 51(1): 34–81.
23. Elhamifar E, Vidal R.Sparse subspace clustering: algorithm, theory, and applications.*IEEE T Pattern Anal Mach Intel*.2013; 35(11): 2765–2781.
24. Smith SM, Brady JM. SUSAN-a new approach to low level image processing. *Int J Comput Vision*.1997; 23(1):45–78.