www.ThePharmaJournal.com

The Pharma Innovation



ISSN (E): 2277-7695 ISSN (P): 2349-8242 NAAS Rating: 5.23 TPI 2023; 12(6): 320-324 © 2023 TPI www.thepharmajournal.com

Received: 15-03-2023 Accepted: 30-05-2023

S Sridevy

Associate Professor, Department of Physical Sciences & Information Technology, Agricultural Engineering College and Research Institute, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

M Nirmala Devi

Assistant Professor, Department of Physical Sciences, Information Technology, Agricultural Engineering College and Research Institute, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

M Sankar

Deputy Librarian, Agricultural College and Research Institute, Tamil Nadu Agricultural University, Tiruvannamalai, Tamil Nadu, India

Nivas Raj Moorthi

Research Scholar, Department of Remote Sensing and GIS, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

Corresponding Author: Nivas Raj Moorthi Research Scholar, Department of

Remote Sensing and GIS, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

Image pre-processing techniques utilized for the plant identification: A review

S Sridevy, M Nirmala Devi, M Sankar and Nivas Raj Moorthi

Abstract

Identification of the plant species is crucial for supporting diversified utilities and applications regarding the biodiversity conservation, Ecological research, taxonomical studies and related systematics, among others. Concerning the difficulties associated with the plant species identification by a non-botanist group, several of the plant species identification framework are being developed and optimized. Lack of an exclusive global consortium group developing the framework might help in increasing the identification accuracies, but are subjected to restrictions concerning computational load associated with the taxonomic complexity, regional variations of the plant species, besides identifying species of convergent evolution. Image processing techniques provides the capabilities for discriminating species through the implication of several learning algorithms coupled with the computer vision. In general, most of the surveyed framework extended their functionalities only at the local or regional level. Further, in regard to the applicability of the framework, most of the system were disseminated as mobile application, for real-time application. This paper provides a comprehensive review on the image pre-processing techniques that can be implicated for the plant species identification.

Keywords: Pre-processing techniques, noise reduction, feature extraction, image segmentation

1. Introduction

Global biodiversity consists of enriched plant richness even among the closely associated plant communities. More than 374,000 accepted number of plant species have been recorded and the discovery of the new species and the updating process are being deteriorated due to the lack of financial support to facilitate the manual species identification techniques. Plant identification can help and support a wide variety of domains including: Biodiversity conservation; Invasive species management; Environmental Impact Assessment; Medicinal and Pharmacological studies; Ecological Research; among others (Christenhusz and Byng, 2016)^[6]. The support can be extended through the global network working on the plant identification studies, and through the implication of the easy access policy of the resource database on request. Plant identification through the manual techniques such as through field guides and floras, Dichotomous Keys, Herbarium specimens, among other includes, destructive sampling for validation, and the process facilitated might be laborious and time-consuming, besides cost expensive (Farnsworth et al., 2013) ^[11]. The central national herbarium at Howrah have collected information of the 2 million specimens and part of which were stored for more than 200 years (Noltie *et al.*, 2021)^[18]. Further, the present methods of plant identification through plant genetic methods, plant chemotaxonomy method, and plant Cytotaxonomy methods also pose a similar disregard to the manual techniques (Wang et al., 2017)^[27]. The lack of trained and skilled taxonomist in examining the specimens collected (taxonomic impediment), also demand the need of exclusive framework, which can integrate the knowledge of researches of all domains in providing an exclusive identification methodology (de Carvalho et al., 2007)^[8]. In general, these methods are based on the subjective Judgement of the plant species identifiers and are subjected to the several limitations. This urged the need of an automated procedure of plant species identification, while training taxonomist in aiding the framework development. With the development of the image processing techniques and through the implication of

frameworks that can handle the computational load, several of the algorithms have been developed and trained in identifying the plant species. With the implementation of the image processing techniques extended to the Medical Imaging (Tumor detection), Surveillance and Security (Face recognition), and forensics (Fingerprint analysis), these applications were considered as the baseline for the image processing-based plant identification procedures (Wang *et al.*, 2017)^[27].

Some of the global initiatives that work on the plant identification and biodiversity conservation includes the Global Biodiversity Information Facility (GBIF) (Edwards, 2004), facilitates sharing of the vast amount of the biodiversity data. In addition, International Barcode of Life (iBOL) project (Adamowicz, 2015)^[1] provides an exclusive DNS barcode reference library for all eukaryotic species. Similarly, other global initiates that have been concentrated on developing a global database of plant species identification includes PlantCLEF (Plant Image Classification at large Scale Evaluation Framework) (Goëau et al., 2022) [12], which compares and validates the image processing techniques for identification at large scale. Flora Incognita Project (Mäder et al., 2021)^[15], combines crowdsourcing and image recognition technologies, aiming at developing a mobile application. In general, the species identification is being facilitated based on the morphometric study of the leaf structure or through the implication of genetic methods of identification including DNA barcoding or PCR (Polymerase Chain Reaction) methods. This review paper provides a scrutiny of the image pre-processing techniques that can be utilized for the plant identification.

With the availability of the digital camera with high specification within our hand-held mobile devices, collection of the specimens can be made instantly and the recommendation of the species identification can be provided instantly. Some of the notable projects include the electronic field guides (Stevenson et al., 2003)^[11] and Leafsnap (Kumar et al., 2012)^[14] of USA, Plant Net (Mutwil et al., 2011)^[17] of France, among others. Of all the features that could be included in training the framework, leaf parameters are considered for their temporal integrity, direct and visual observation of the leaf structural parameters and its availability in most of the seasons, For the development of the framework, collaboration from both the botanist and the computer scientist is required (Wäldchen et al., 2018)^[25, 26]. Some of the challenges that were specified in developing a digital morphometrics system of plants by Cope et al. (2012) ^[7] includes, 1. 3- Dimensional nature of the leave and the mechanical damage from the external stresses can degrade the potential of the images utilized for identification; 2. State of confusion between the classification and clustering terminologies associated with the botanist/taxonomist and computer scientist, respectively; 3. Variation in the leaf structure boundaries from a single plant; 4. Confluence of the multiple feature information for increasing the accuracy of the results - resulting in higher computational load; 5. Other

ambiguities associated with the camera parameters, selecting appropriate algorithms, illumination constrains might also degrade the potential of the identification results. The objective of this review involves studying different leaf structure input parameters and stating the different classifiers that have been incorporated in several of the studies.

2. Typical framework of the plant species identification

The typical framework of the plant species identification has been depicted in the Figure 1. The major phases, as depicted in the figure 1, includes image collection and preprocessing, pattern recognition and image classification (Wäldchen and Mäder, 2018)^[25, 26]. In general, the image obtained from the camera devices must be of the specified resolution, as some of the enhancement techniques alters the dimensions of the pixel cells if required, ultimately changing the resolution. As most of the image pre-processing techniques works on the frequency domain of the image rather than the spatial domain, the originality of the images can be retained utmost. The image processing techniques that are commonly employed are to reduce the effects of differing illumination (affecting the contrast), interference and noise removal, which initial converts the colored image to gray scale, as most of the studies discarded the color parameter for feature extraction. Before the feature extraction, pattern recognition and classification process, it is evident that the images must be subjected to the pre-processing. Preprocessing generally affects the quality and the quantity of the input datasets, primarily affects the results of the recognition process. The feature extraction process generally includes the extracting of the image features indicating the morphometric parameters, which are to be considered as the independent variable, based on which the pattern recognition and classification process is implemented. The leaf morphometric features that considered for extraction includes leaf shape (shape descriptors), leaf textural analysis (statistical measures) leaf vein structure and arrangement, leaf marginal analysis, and color information (RGB/HIS), of which most of the framework adopted leaf shape analysis parameters for their framework analysis. The automatic extraction of the intricate parameters through the leaf shape analysis after preprocessing is considered essential for the subsequent analysis. All the features or parameters extracted were converted to the vector format, which will be utilized by the learning algorithms for classification and providing the final results. The scope of this review identifies and compares the image pre-processing techniques that can be utilized for the plant species identification.



Fig 1: Framework of plant species identification

3. Image processing

Image processing techniques determines the quality and the quantity of the raw image datasets that are to incorporated for the subsequent analysis. In general, the image processing techniques are employed from the image pre-processing step to that of the feature extraction step (Dhaware and Wanjale, 2017). From the literature survey, image pre-processing techniques that have been identified have been are discussed in this following subsection. Concerning the image processing techniques that are capable of improving the quality of the image for identification, the major processes that were included in the discrimination includes, image normalization, image segmentation, dimensionality reduction and image recognition primarily based on the leaf morphometric studies, which requires the users to take digital images at various angles at specified resolution.

3.1 Image pre-processing techniques

In general, the selection of the pre-processing techniques depends on the learning algorithm requirements or depends upon the quality of the raw digital images captured. In both the context, the applicability of the pre-processing techniques is indispensable considering the differing illumination perspective of the image, besides the noise and the interference associated with the image during image capture. Further, considering the larger dataset requirement for training the algorithm, the redundancy associated with the raw images must be accounted when subjecting the images for feature extraction. Scrutiny of the image pre-processing techniques must be implemented in regard of considering the originality of the raw images, as some of the techniques might alter the image in both the frequency and the spatial domain and has overlapping advantages and limitation. The following subsection discusses the image pre-processing techniques that can be considered for the subsequent feature extraction and classification techniques. The mentioned pre-processing techniques can be implemented sequential or can be combined in any order, which are a function of the image acquisition parameters and the features to be extracted. Rzanny et al. (2017) ^[23] compared the effective of three different preprocessing strategies (i.e.) Non-preprocessed, Cropped and segmentation for the automated plant species identification and the image features are acquired through the CNN, and the study concluded that pre-processing measures does not have any substantial increase in the functioning of the framwork.

3.1.1 Image Filtering

Image Filtering involves implication of the filter or kernel of certain window size to perform operations such as smoothening, sharpening, noise reduction and edge detection, among others. The majorly employed filtering techniques includes, Gaussian Filter, Mean Filter, Median Filter, Bilateral Filter, Laplacian Filter, Sobel Filter, Canny Edge Detection, High-pass and low pass filters, Anisotropic Diffusion, each of which has its own gain and loss in the image spectrum (Chandel and Gupta, 2013)^[5]. The major implication of performing filtering on the input raw datasets is to remove the noise associated with the images leading to the derivation of visually appealing images. The other advantages that the filtering techniques includes smoothening and blurring of the irrelevant information and can help in identify/highlight boundaries (Sobel operator) of different regions or objects. The boundary identification can help in the subsequent process of image segmentation and feature extraction. Ying et al. (2009)^[30] compared simple filter and mean filter in detecting the cucumber powdery mildew diseases and stated that the image preprocessing techniques increased the effective of the classification results. The challenges or the limitations that were imparted during filtering includes, selection of a specific filtering algorithms that balances the trade off between the noise reduction and detail preservation of the images, computational complexity associated with the anisotropic or edge detection algorithms, appropriate selection of the filter parameters and unlikely selection of filter parameter might lead to information loss and can introduce blurring, or strips or halos around the edges of the image.

3.1.2 Image resizing/rescaling

For enabling the plant species identification, often the

acquired raw images must be of same proposed dimensions. Though rescaling and resizing are very closely related image processing techniques, image rescaling techniques aims at changing the dimensions of the image while maintaining the original aspect ratio and image resizing aims at changing the dimensions, with or without maintaining the aspect ratio of the image. Some of the image resizing/ rescaling techniques that can be considered includes, proportional resizing; resizing; bilinear interpolation; absolute Lanczos interpolation; Nearest neighbour interpolation; bicubic interpolation; Seam carving and super resolution. Each of these techniques provides effective resizing/rescaling of the images, offering its own advantages and limitations (Parsania and Virparia, 2014)^[19]. Some of the classification algorithms particularly of deep learning architectures, or stacking of the input raw images requires the images to be of standard dimensions or has input size constraints, hence resizing techniques helps in the resizing the images to the requirements. Further, resizing helps in the increasing the computational efficiency and reduces the noise associated with the images to a certain extend. Resizing also optimizes the storage facilitations and provides enhanced visualization of the images by truncating the unnecessary features of the image, providing grounds for manual identification. In general, most of the studies employed resizing dimensions that maintain the original aspect ratio ensure that the proportions of the plant objects are maintained, facilitating accurate feature extraction and classification. Resizing also poses some of the limitations. Resizing will lead to loss of important information if the aspect ratio of the original images is not preserved. This might also introduce distortions and artifacts such as blurring, jagged edges or pixelation. In addition, the initial quality of the images must be considered, as resizing low quality images might amplify noise, artifacts and other aberrations in the intensity values of the image. The image resizing techniques have been implemented in several of the plant identification studies (Wei Tan et al., 2018; Kaur and Kaur, 2019; Pereira et al., 2019)^[28, 13, 20].

3.1.3 Image Normalization

Image normalization techniques are used to standardize the pixel values or characteristics of an image. These techniques help in reducing variations in image data, enhancing the comparability of images, and improving the performance of subsequent image processing or analysis tasks. Some of the image normalization techniques include, Minimum-Maximum Normalization, Z- Score Normalization, Contrast stretching, Histogram Equalization, log transformation, power -law (Gamma) Correction, among others. In general, image normalization enhances the contrast of the image by implementing the above-mentioned techniques. Similar to the other pre-processing steps, image normalization techniques also provide facilities for increasing the image quality for comparison and visualization, can normalize the pixel values to the standardized range, such as [0, 1] or [-1, 1], normalization helps in achieving numerical stability, convergence, and improved performance of machine learning algorithms. It prevents features with larger magnitudes from dominating the learning process and ensures that all features contribute equally (Prasad et al., 2011)^[21]. Normalization techniques preserve these relative relationships while adjusting the overall scale or distribution of pixel values. This ensures that important image features or patterns are retained even after normalization. By reducing noise, enhancing contrast, and highlighting relevant image features, normalization helps in better segmenting objects, detecting edges, extracting meaningful features, and performing subsequent analysis, such as classification or recognition (Woebbecke *et al.*, 1993)^[29]. The major limitation of the normalization techniques is the introduction of subjectivity and potential bias into the analysis or interpretation of images, when applying a normalization technique that is not suited or required for the particular task. In addition, normalization techniques pose limitations from the loss of information, artefact introduction, parameter dependency and over-amplification of the noise in the images.

3.1.4 Image Segmentation

Image segmentation is an important process of the plant identification framework, as it forms the basis for the further feature extraction and pattern recognition. It involves partitioning an image into meaningful and distinct regions or objects. The goal of image segmentation is to extract and separate different regions of interest based on their visual characteristics, such as color, texture, intensity, or spatial relationships (Meyer et al., 1999) ^[16]. The resulting segmentation can facilitate further analysis, object recognition, tracking, or understanding of image content. Image segmentation techniques can be broadly categorized into the following approaches such as Thresholding (setting an threshold values and classifying the pixels based on their intensity values), Region based segmentation (grouping pixels together based on certain similarity criteria to form meaningful regions or objects), Edge-Based Segmentation (Canny edge detector or the Sobel operator can extract edges, which can then be used to delineate object boundaries), Clustering (grouping pixels based on similarity in feature space), Graph-Based Segmentation (models the image as a graph, where pixels are represented as nodes, and the relationships between pixels are defined by edges), Deep Learning-Based Segmentation (Fully Convolutional Networks (FCNs), U-Net, and Mask R-CNN are popular architectures for semantic segmentation, instance segmentation, and pixellevel segmentation tasks) (Ambarwari et al., 2020)^[13].

3.1.5 Dimensionality Reduction

Considering the large number of images taken at different angles, redundancy is often associated with the image datasets. To reduce the complexity and dimensionality of the feature space and to provide true and inherent image data sets for feature extractions, dimensionality reduction aims to extract the most informative features or transform the data into a lower-dimensional representation while preserving or maximizing the discriminative information (Ahmed et al., 2016) ^[2]. Some of the measures includes, Principal Component Analysis (PCA), Linear Discriminant Analysis, t-Distributed Stochastic Neighbor Embedding (t-SNE), Isomap, Autoencoders, and Random Projection. Quoc Bao et al. (2020) ^[22] proposed a framework for plant species identification from the leaf patterns, which utilized BoW model for the dimensionality reduction. In some cases, dimensionality reduction is applied before image segmentation as a preprocessing step. This is done to reduce the complexity and dimensionality of the feature space before segmenting the image into meaningful regions or objects. By reducing the dimensionality, the segmentation algorithm can

operate on a lower-dimensional feature space, which can lead to improved efficiency and performance. Dimensionality reduction techniques like PCA or LDA can be applied to extract the most informative features from the original image data before segmentation. Alternatively, dimensionality reduction can also be applied after image segmentation. Once the image is segmented into distinct regions or objects, dimensionality reduction techniques can be used to extract lower-dimensional representations of each segmented region. This can be done to reduce the dimensionality of the feature vectors describing each region, making subsequent analysis or classification tasks more efficient. Dimensionality reduction techniques like PCA or LDA can be applied to the feature vectors extracted from each segmented region. The choice of when to apply dimensionality reduction depends on the specific requirements and characteristics of the plant identification task. If the dimensionality of the original image data is high and needs to be reduced for computational efficiency or improved performance, applying dimensionality reduction before segmentation may be beneficial. On the other hand, if the focus is on reducing the dimensionality of the feature vectors extracted from segmented regions, dimensionality reduction can be applied after segmentation (Bingham et al., 2006)^[4].

4. Conclusion

In conclusion, pre-processing techniques play a vital role in plant species identification using image processing. These techniques help enhance the quality and suitability of input images, making them more suitable for subsequent analysis and classification. By applying various pre-processing steps such as image enhancement, noise reduction, image resizing, and color normalization, the images can be optimized to reveal essential features and reduce unwanted variations. It aids in improving image quality, enhancing contrast, and reducing noise, thereby increasing the effectiveness of subsequent analysis algorithms. Pre-processing techniques enable the extraction of meaningful features, such as shape, texture, and color, from the plant images, which are crucial for distinguishing different species.

Additionally, pre-processing techniques help overcome challenges related to varying lighting conditions, image artifacts, and inconsistencies in image acquisition. By standardizing the images and removing unwanted artifacts, the subsequent identification algorithms can work more effectively and produce more reliable results. The combination of pre-processing techniques with advanced image processing algorithms, machine learning, and pattern recognition techniques has revolutionized the field of plant species identification. These techniques have opened up new possibilities for automated, rapid, and accurate identification of plant species, benefiting various domains such as agriculture, ecology, conservation, and botany. However, it is important to note that the effectiveness of pre-processing techniques and subsequent identification algorithms depends on the quality of the input images and the complexity of the plant species. Some plant species may pose challenges due to variations in leaf shape, color, and texture, which can impact the accuracy of the identification process. Therefore, continuous research and advancements in pre-processing techniques, image analysis algorithms, and machine learning models are essential to further improve the accuracy and robustness of plant species identification systems.

References

- 1. Adamowicz SJ. International Barcode of Life: Evolution of a global research community. Genome. 2015;58(5):151-162.
- Ahmed N, Khan UG, Asif S. An Automatic Leaf Based Plant Identification System. Science international, 2016, 28(1).
- 3. Ambarwari A, Adrian QJ, Herdiyeni Y, Hermadi I. Plant species identification based on leaf venation features using SVM. Telkomnika Telecommunication Computing Electronics and Control. 2020;18(2):726-732.
- Bingham E, Gionis A, Haiminen N, Hiisilä H, Mannila H, Terzi E. Segmentation and dimensionality reduction; c2006.
- 5. Chandel R, Gupta G. Image filtering algorithms and techniques: A review. International Journal of Advanced Research in Computer Science and Software Engineering, 2013, 3(10).
- 6. Christenhusz MJ, Byng JW. The number of known plants species in the world and its annual increase. Phytotaxa. 2016;261(3):201-217-201-217.
- Cope JS, Corney D, Clark JY, Remagnino P, Wilkin P. Plant species identification using digital morphometrics: A review. Expert Systems with Applications. 2012;39(8):7562-7573.
- 8. de Carvalho MR, Bockmann FA, Amorim DS, Brandão CRF, de Vivo M, de Figueiredo JL, *et al.* Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cyber taxonomic-automation paradigm. Evolutionary Biology. 2007;34:140-143.
- 9. Dhaware CG, Wanjale K. A modern approach for plant leaf disease classification which depends on leaf image processing. International Conference on Computer Communication and Informatics (ICCCI); c2017.
- Edwards JL. Research and societal benefits of the Global Biodiversity Information Facility. Bio Science. 2004;54(6):485-486.
- 11. Farnsworth EJ, Chu M, Kress WJ, Neill AK, Best JH, Pickering J, Stevenson RD, *et al.* Next-generation field guides. Bio Science. 2013;63(11):891-899.
- 12. Goëau H, Bonnet P, Joly A. Overview of PlantCLEF 2022: Image-based plant identification at global scale. CLEF 2022-Conference and Labs of the Evaluation Forum; c2022.
- Kaur S, Kaur P. Plant species identification based on plant leaf using computer vision and machine learning techniques. Journal of Multimedia Information System. 2019;6(2):49-60.
- Kumar N, Belhumeur PN, Biswas A, Jacobs DW, Kress WJ, Lopez IC, *et al.* Leafsnap: A computer vision system for automatic plant species identification. Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy; c2012 October. p. 7-13. Proceedings, Part II 12.
- 15. Mäder P, Boho D, Rzanny M, Seeland M, Wittich HC, Deggelmann A, *et al.* The Flora Incognita app–interactive plant species identification. Methods in Ecology and Evolution. 2021;12(7):1335-1342.
- Meyer GE, Hindman TW, Laksmi K. Machine vision detection parameters for plant species identification. Precision agriculture and biological quality; c1999.
- 17. Mutwil M, Klie S, Tohge T, Giorgi FM, Wilkins O, Campbell MM, et al. Pla Net: combined sequence and

expression comparisons across plant networks derived from seven species. The Plant Cell. 2011;23(3):895-910.

- Noltie HJ, Kumar A, Bharati KA, Bhattacharjee A, Krishna G. Specimens of William Roxburgh in the Central National Herbarium at the AJC Bose Indian Botanic Garden, Howrah. Current science. 2021;120(6):997.
- 19. Parsania P, Virparia PV. A review: Image interpolation techniques for image scaling. International Journal of Innovative Research in Computer and Communication Engineering. 2014;2(12):7409-7414.
- 20. Pereira CS, Morais R, Reis MJ. Deep learning techniques for grape plant species identification in natural images. Sensors. 2019;19(22):4850.
- 21. Prasad S, Kumar P, Tripathi R. Plant leaf species identification using curvelet transform. 2011 2nd International Conference on Computer and Communication Technology (ICCCT-2011); c2011.
- 22. Quoc Bao T, Tan Kiet NT, Quoc Dinh T, Hiep HX. "Plant species identification from leaf patterns using histogram of oriented gradients feature space and convolution neural networks. Journal of Information and Telecommunication. 2020;4(2):140-150.
- 23. Rzanny M, Seeland M, Wäldchen J, Mäder P. Acquiring and preprocessing leaf images for automated plant identification: understanding the tradeoff between effort and information gain. Plant Methods. 2017;13(1):97. DOI: 10.1186/s13007-017-0245-8.
- 24. Stevenson RD, Haber WA, Morris RA. Electronic field guides and user communities in the eco-informatics revolution. Conservation Ecology, 2003, 7(1).
- Wäldchen J, Mäder P. Plant species identification using computer vision techniques: A systematic literature review. Archives of Computational Methods in Engineering. 2018;25:507-543.
- Wäldchen J, Rzanny M, Seeland M, Mäder P. Automated plant species identification-Trends and future directions. PLoS computational biology. 2018;14(4):e1005993.
- 27. Wang Z, Li H, Zhu Y, Xu T. Review of plant identification based on image processing. Archives of Computational Methods in Engineering. 2017;24:637-654.
- Wei Tan J, Chang SW, Abdul-Kareem S, Yap HJ, Yong KT. Deep learning for plant species classification using leaf vein morphometric. IEEE/ACM transactions on computational biology and bioinformatics. 2018;17(1):82-90.
- 29. Woebbecke DM, Meyer GE, Von Bargen K, Mortensen DA. Plant species identification, size, and enumeration using machine vision techniques on near-binary images. Optics in Agriculture and Forestry; c1993.
- Ying G, Miao L, Yuan Y, Zelin H. A Study on the Method of Image Pre-processing for Recognition of Crop Diseases. 2009 International Conference on Advanced Computer Control; c2009.