3RD INTERNATIONAL CONFERENCE ON INNOVATIVE COMPUTING AND COMMUNICATION Indian ayurvedic plant identification using multi-organ image analytics: creation of image dataset of Indian...

Conference Paper · March 2020

DOI: 10.2139/ssrn.3563074.

2 authors:

Meera Kansara
Gujarat Vidyapith

Ajay Dineshchandra Parikh
Gujarat Vidyapith

Indian ayurvedic plant identification using multi-organ image analytics: creation of image dataset of Indian medicinal plant organs

Meera Kansara[a], Ajay Parikh[a]*

[a]*Department of Computer Science, Gujarat Vidyapith, Ahmedabad 14, Gujarat, India*

Abstract: This paper focuses on process standardization for creation of image dataset of Indian medicinal plant organs. There are factors like climate, time of year, geographical distribution, soil condition and biodiversity that can affect plants. Based on the standardized process we have prototyped mobile based tool that helps to acquire lots of structured plant organ images. Along with images tool captures geographic location and time of year which aids in inferencing plant species that can occur at a certain geographic location during certain time of year. For dataset creation this research considered aspects like time of year, time of day, illumination conditions and backgrounds. The interfaces utility takes care of primary health check and outlier detection.

Keywords: Indian medicinal plant identification, Leaf, Flower, Plant multi-organ image dataset, Mobile based image capturing tool, Image analytics.

1. Introduction

Medicinal Plants form the primary resource base of Indian autochthonous health care traditions. Almost decade ago (Gaston & O'Neill, 2004) proposed feasibility of species identification using artificial intelligence and digital image processing techniques. Since then number of studies have proposed various methods for automated Indian ayurvedic medicinal plant identification. (Rzanny, Seeland, Wäldchen, & Mäder, 2017a) explored several methods of image acquisition and pre-processing to enhance the quality of plant organ images to train classifiers for classification.

This paper proposes image dataset dedicated to Indian ayurvedic medicinal plant organs to carry out an image-based plant identification and classification. To aid image acquisition task and standardize parameters for image acquisition, we developed mobile based image capturing tool. This tool is convenient and encourages non-destructive way to capture plant organ images. The mobile based image capturing can be used by non-expert user and by following basic guidelines one can capture images of plant organ with ease. The other important tool is gateway interface which manages dataset and responsible for primary health check of dataset. Gateway interface has utility for outlier detection which identifies outlier or anomalies introduced during image acquisition process. This work is exclusive as there is no standard plant organ image dataset for Indian ayurvedic medicinal plants. The proposed dataset considers image acquisition from various aspects like time of day, time of year, illumination condition and background. The supplementary information such as geographic location, climatic condition can be very useful for real time plant identification.

2. Related work

Many efficient and innovative approaches have been proposed in recent time for automatic identification of Indian ayurvedic medicinal plants. Table 1 summarizes the researches carried out in recent time. (Sabu, Sreekumar & Nair, 2017) used HoG (Histogram of oriented Gradients) and SURF (Speeded Up Robust Features) features along with k-NN classifier to identify ayurvedic medicinal plants. There are three approaches ((Vijayashree & Gopal, 2017), (Pushpa, Anand & Nambiar, 2016), (Kumar & Talasila, 2014)) which did not used state of the art classifiers but instead used dissimilarity calculation (Vijayashree & Gopal, 2017), unique statistical factor called leaf factor (Pushpa, Anand & Nambiar, 2016), unique identifier (Kumar & Talasila, 2014) for leaf and carried out classification. Another work was carried out by (Kumar, Surya & Gopi, 2017), the research considered both front and back side of leaves with fresh and dried leaves, features were extracted and tested with classifiers including Support Vector Machine (SVM) and Multi-Layered Perceptron (MLP). Closely related work was done by (Dahigaonkar & Kalyane, 2018) by extracting various geometric, texture, shape and color features and SVM Classifier. (Venkataraman & Mangayarkarasi, 2017) used various statistical parameters, texture features and SVM for classification and identification of Indian medicinal plants. Artificial Neural Network (ANN) and Convolution Neural Network (CNN) were respectively used by (Aitwadkar, Deshpande & Savant, 2018) and (Batvia, Patel & Vasant, 2017) for automatic identification of Indian medicinal plants. Another considerable work was carried out by (Venkataraman & Mangayarkarasi, 2016) using different sized leaves of the same plant to analyze which features vary with size and which features are constant and concluded that aspect ratio remains constant. (Arun & Christopher Durairaj, 2017) used Grey Tone Spatial Dependency Matrix (GTSDM), Local Binary Pattern (LBP) based features with 20 different color spaces and proved better identification with HSV and YUV, L*a*b*. The following Table 1 gives a summary of recent work in the field of automatic Indian ayurvedic medicinal plant identification.

Table 1 - Recent work in the area of automatic Indian ayurvedic medicinal plant identification.

Ref.	Area of study	Primar y feature	No of sam	Plan t orga	classifier	dataset

		s	ples	n		
(Sabu, Sreekumar & Nair, 2017)	India	SURF, HOG	200	Leaf	k-NN	Custom
(Vijayashree & Gopal, 2017)	India	Texture	127	Leaf	Dissimilarity	Custom
(Pushpa, Anand & Nambiar, 2016)	India	Shape	208	Leaf	Leaf factor	Custom
(Kumar & Talasila, 2014)	India	Shape, Texture, Color	500	Leaf	Unique ID	Custom
(Kumar, Surya & Gopi, 2017)	India	Color, Texture, Geometric CR distance	1200	Leaf	SVM, MLP, and others	Custom
(Dahigaonkar & Kalyane, 2018)	India	Shape, Color, Texture, Geometric	128	Leaf	SVM	Custom
(Venkataraman & Mangayarkarasi, 2017)	India	Texture	260	Leaf	SVM	Custom
(Aitwadkar, Deshpande & Savant, 2018)	India	Edge, Color	50	Leaf	ANN	Custom
(Batvia, Patel & Vasant, 2017)	India	-	4000 (approx)	Leaf	CNN	Custom
(Venkataraman & Mangayarkarasi, 2016)	India	Shape	5	Leaf	ANN, SVM	Custom
(Arun & Christopher Durairaj, 2017)	India	Color, Texture	250	Leaf	SGD, SVM with RBF kernel	Custom

From the details presented in Table 1 two points are noticeable, First, researches in the field of Indian ayurvedic medicinal plant identification mainly focuses on single plant organ (leaf). Second, the researchers are creating custom dataset for their research as there is no standard dataset available for Indian medicinal plant organs. Below mentioned Table 2 enlists some of the existing plant image datasets used by the researcher in the field of automatic plant identification/classification.

Table 2 – Some of the existing plant image datasets.

Dataset	Region/ Country	Organ	No. of Species	No. of Images
Swedish Leaf (Söderkvist, O., 2001)	Sweden	Leaf	15	1125
Flavia (Wu, et al., 2007)	China	Leaf	32	1907
Leafsnap (Kumar, et al., 2012)	North-eastern United States	Leaf	184	30866
Malaya Kew Leaf (Lee, Chan, Wilkin & Remagnino, 2015)	England	Leaf	44	2288
Oxford flower 102 (Nilsback & Zisserman, 2008)	United Kingdom	Flower	102	8189
Oxford flower 17 (Nilsback & Zisserman, 2006)	United Kingdom	Flower	17	1360
Jena Leaf Images 17 (Rzanny, Seeland, Wäldchen, & Mäder, 2017b)	Jena, Germany	Leaf	17	2902

The key point to note in the Table 2 is that none of above plant organ image datasets are dedicated to Indian ayurvedic medicinal plant organs. This research addresses the need for benchmark dataset for Indian medicinal plant organs.

3. Proposed methodology

The primary need for accurate plant identification is standard dataset of plant organ images. The dataset creation pipeline consists of several stages as described in Fig. 1 below.

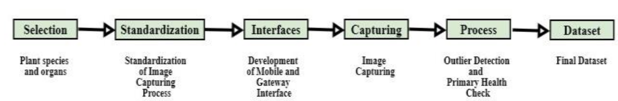

Fig. 1 – Dataset creation pipeline.

3.1. Plant selection

Selection of plant based on geographical area and medicinal values. For this research we are considering 50 medicinal plant species (Jadeja et al., 2006).

3.2. Standardization

Process Standardization includes standardization of dataset structure, naming convention and parameters (like capturing time of year, geographic location). The detailed discussion of hierarchical structure and naming convention is as follows.

Dataset structure

Following are major aspects this research considered to create dataset of plant organ images:

- Plant organ (leaf, flower).
- Time of Year-Image capturing on monthly basis.
- Time of Day-Image capturing at different time of day.
- Illumination-Image capturing with different illumination like direct sunlight or in shadow, flash on or flash off.
- Background-Image capturing in natural environment or in lab settings where background is fixed color sheet and scanned images.

In the proposed dataset, for every plant species and plant organ five types of images will be captured as follows (Fig. 2).

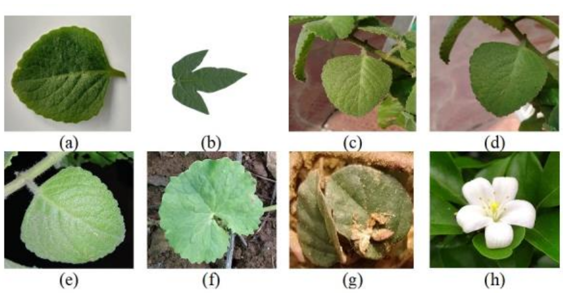

Fig. 2 – Sample images of aspects (a) leaf-lab image, (b) leaf-scanned image, (c) leaf-morning-natural background-direct sunlight, (d) leaf-morning-natural background- shadow, (e) leaf-night-natural background-flash on, (f) leaf-evening-natural background-without flash, (g) leaf-occulted image, (h) flower-morning-natural background-direct sunlight.

1. Lab images which consist of month wise scan and in-lab image capturing. In-lab are the images taken with pre-defined static background and static standard illumination (Fig. 2(a)). Scanned Images simply means putting leaf on scanner and scanning the plant organs (Fig. 2(b)). This is done monthly bases for each plant species and organ.

2. Field images are images taken in natural background monthly and during different time of day, Morning, Afternoon, Evening and Night (Fig. 2(c, d, e, f, h)). In the morning and afternoon images are captured in two illumination condition viz. direct sunlight and in shadow. In the evening and night images are captured with two illumination conditions viz. flash on and flash off.

3. Occulted images, an imperfect images of plant organ, this includes pictures of broken or imperfect organs which are cut off from view due to various reason and are not captured fully (Fig. 2(g)).

4. Perfect Images, A high resolution images of plant organ taken by professional image capturing device and edited perfectly.

5. Microscopic images are as name suggest Microscopic images of the organ.

In dataset structure (Fig. 3), plant scientific name is at the root of the hierarchy and followed by hierarchical structure. Similar kind of image capturing will be carried out for all plant organs depending upon their availability during the year.

Naming Convention

For accuracy and consistency dataset will follow definite structure and strict naming convention where parameter information is encoded within each image name. The image name is primarily divided in to three parts: Plant scientific name, encoded tree structure and sequence number.

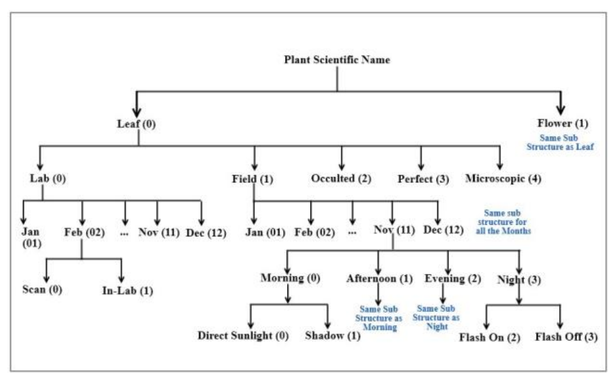

Fig. 3 - Hierarchical structure- plant organ image dataset.

The encoded part contains information about plant organ, image background or type, time of year, time of day, illumination conditions (Table 3). For Example. Image name, *Aegle marmelos_010700_1111*.

Table 3 – Image naming convention encoding.

Aspects	Encoding	Example encoding

Organ	0	Leaf	First digit 0 means Leaf
	1	Flower	
Image Background and Type	0	Lab	Second single digit 1 means it is image taken in field (natural background)
	1	Fields	
	2	Occulted	
	3	Perfect	
	4	Microscopic	
Time of Year (month)	01	January…	Third-Fourth two digits denote month, 07 means July
	12	December	
Time of Day	0	Morning	Fifth digit denote Time of day, 0 means image taken in the morning
	1	Afternoon	
	2	Evening	
	3	Night	
Illumination Condition	0	Direct sunlight	Sixth digit denote Illumination conditions, 0 means image taken in direct sunlight.
	1	Shadow	
	2	Flash On	
	3	Flash Off	

The first component is plant scientific name *Aegle marmelos*, second component is encoded part *010700* indicates leaf image taken in the natural background during month July in the morning under direct sunlight (Table 3). The third component *1111* is sequence number.

3.3. Interfaces (Development of image acquisition tools)

Image acquisition is a complex and time-consuming task. In image-based classification amount and quality of images plays an important role in achieving higher accuracy. To address this issue, we have created utilities to carry out efficient image capturing and administration.

Image capturing mobile tool and Image gateway interface

The image capturing mobile tool is developed for Android mobile phones to capture plant organ images. The mobile capturing tool helps user to adhere to standardized structure by providing visual guidelines for accurate image capturing. This reduce chance of capturing highly noised images or inaccurate images. Hence contributes in creating dataset which may requires less pre-processing. The information like the geographic location and time of year gives valuable background information which can expediate plant species identification. Image gateway is an administrative interface which serves as Gateway of dataset and controls addition of images based on parameters in dataset. Apart from dataset administration, this interface is responsible for outlier detection and treatment of outlier in dataset.

3.4. Image capturing

The step where actual image capturing is being carried out using image capturing mobile tool and considering the aspects discussed in Standardization section (Section 3.2) above.

3.5. Process (Outlier detection)

Outlier detection safeguards against accidental miss-labelled images, wrongly captured image, images that are not in line with guidelines and offers first level health check for dataset. In process first step is selection of images for which outlier detection has to be carried out. The pre-processing step includes operation such as normalization, smoothing. The outlier detection step carries out actual outlier detection and images of outlier plant organ will get displayed for subsequent action.

Fig. 4 – Sample raw images.

To carry out an elementary health checkup and probable outlier detection we captured images of *Coleus amboinicus* along with images of other plants (Fig.4), then extracted various Shape, Color and Texture features from each image. In total 17 features were extracted. We used state of the art Isolation Forest algorithm with extracted feature values to detect possible outlier in image dataset. To implement this utility, we used OpenCV (version 4.1.0) an open source computer vision library with Python (Bradski, 2000) to extract Shape features, Color features and Texture features and Isolation Forest algorithm of scikit-learn python library (version 0.21.2) (Buitinck et al., 2013).

Along with outlier detection this health check utility also enlists plant organ images that does not adhere to guidelines like images missing geo location information.

3.6. Dataset creation

The stable final dataset is created after execution of all above mentioned processing steps and will be available publicly to be used by researchers.

4. Discussion

Following comparison table (Table 4) enlists the benefits of using mobile based image capturing tool compared to direct image capturing in context of this research.

Table 4 – Comparison of direct and tool-based image capturing.

Aspects	Direct image capturing	Image capturing using mobile tool	Impact
Geo location enforcement	Difficult to enforce	Yes	Geo tagging and inference based on geo location information.
User identification	Difficult to achieve	Yes	user efficiency in terms of number of, quality and quantity of samples collected/ training required.
Consistency	No (across multiple devices)	Yes	consistent image quality in terms of size and shape.
Plant identifier	No (unless user save each plant images in separate directory manually)	Yes	Accurate and clean dataset.
Detailed aspect information (type or organ, time of year, time of day, illumination, background)	No	Yes	This information can be very useful for identification and inferencing, also used for comparative analysis.
Validation and enforcement of aspects	No	Yes	Clean, consistent and more stable dataset.
Auto sync with repository without manual intervention	No	Yes	Up to date information and less prone to loss of data due to device failure or accidental removal.
Outliers	Can be more as images do not have labels assigned while capturing	Comparatively less	Less erroneous images and clean dataset.
Manual efforts	More	Comparatively less	Saves associated efforts, cost and time.

For primary experiment of outlier detection, we compared algorithms Isolation Forest and Local Outlier Factor with the 17 features extracted from each image of the sample dataset (Fig. 4).

Table 5 –Results of outlier detection for sample image dataset.

Method	Accuracy	Precision	Recall	F1 score
Isolation Forest	88.79%	64.01%	80.00%	71.11%
Local Outlier Factor	86.20%	57.69%	75.01%	65.21%

As shown in Table 5 above, Isolation forest achieved 88.79% and Local outlier factor achieved 86.20% accuracy respectively. Isolation forest performed better in comparison to Local outlier factor with extracted feature set and sample dataset. The results can be further improved by using larger sample dataset and extracting more efficient features.

5. Conclusion

We considered creation of the Indian medicinal plant organ image dataset, with structured approach and different aspects of image acquisition. The mobile based tool for image acquisition helps in capturing structured images and reducing efforts related to data cleaning. We also carried out primary experiments for outlier detection and found Isolation Forest giving more accurate results. The mobile based tool and dataset will be outcome of the first iteration of Indian ayurvedic plant identification using multi-organ image analytics. Dataset will be released to be used by researchers to save their time, efforts and cost associated with dataset creation.

Acknowledgements

We would like to thank Dr. Minubhai Parabia, Dr. Srinivas Murty and Mr. Jignesh Kansara for their valuable comments on an earlier version of the manuscript.

References

Gaston, K. J., & O'Neill, M. A. (2004). Automated species identification: why not?. Phil. Trans. R. Soc. Lond. B,359, 655-667. http://doi.org/10.1098/rstb.2003.1442

Rzanny, M., Seeland, M., Wäldchen, J., & Mäder, P. (2017a). Acquiring and preprocessing leaf images for automated plant identification: understanding the tradeoff between effort and information gain. Plant Methods 13, 97. doi:10.1186/s13007-017-0245-8

Sabu, A., Sreekumar, K., & Nair, R. (2017). Recognition of ayurvedic medicinal plants from leaves: A computer vision approach. Fourth International Conference on Image Information Processing, 574-578.

Vijayashree, T., & Gopal, A. (2017). Leaf identification for the extraction of medicinal qualities using image processing algorithm. International Conference on Intelligent Computing and Control, Coimbatore, 1-4. doi:10.1109/I2C2.2017.8321884

Pushpa, B. R., Anand.,C., & Nambiar Mithun, P. (2016). Ayurvedic Plant Species Recognition using Statistical Parameters on Leaf Images. International Journal of Applied Engineering Research, 11(7), 5142-5147

Kumar, S. E. & Talasila, V. (2014). Leaf features based approach for automated identification of medicinal plants. International Conference on Communication and Signal Processing, 210-214. doi:10.1109/ICCSP.2014.6949830

Kumar, M. P., Surya, C. M., & Gopi, V. P. (2017). Identification of ayurvedic medicinal plants by image processing of leaf samples. Third International Conference on Research in Computational Intelligence and Communication Networks, 231-238.

Dahigaonkar, T., & Kalyane, R. (2018). Identification of Ayurvedic Medicinal Plants by Image Processing of leaf samples. International Research Journal of Engineering and Technology, 5(5), 351-355.

Venkataraman, D., & Mangayarkarasi, N. (2017). Support vector machine based classification of medicinal plants using leaf features. International Conference on Advances in Computing, Communications and Informatics, 793-798.

Aitwadkar, P. P., Deshpande, S. C., & Savant, A.V. (2018). Identification of Indian medicinal plant by using artificial neural network. International Research Journal of Engineering and Technology, 5(4),1669–1671.

Batvia, V., Patel, D., & Vasant, A. R. (2017). A Survey on Ayurvedic Medicine Classification using Tensor flow. International Journal of Computer Trends and Technology, 53(2), 68-70.

Venkataraman, D., & Mangayarkarasi, N. (2016). Computer vision based feature extraction of leaves for identification of medicinal values of plants. IEEE International Conference on Computational Intelligence and Computing Research, 1-5.

Arun, C., & Christopher Durairaj, D. (2017). Identifying Medicinal Plant Leaves using Textures and Optimal Colour Spaces Channel. Jurnal Ilmu Komputer dan Informasi, 10(1), 19-28. doi:http://dx.doi.org/10.21609/jiki.v10i1.405

Söderkvist, O. (2001). Computer Vision Classification of Leaves from Swedish Trees (Dissertation). <http://www.cvl.isy.liu.se/en/research/datasets/swedish-leaf/> (last accessed 2019/04/11).

Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y., Chang, Y., & Xiang, Q. (2007). A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network. IEEE International Symposium on Signal Processing and Information Technology, Giza, 11-16. doi: 10.1109/ISSPIT.2007.4458016

Kumar, N., Belhumeur, P. N., Biswas, A., Jacobs, D. W., Kress, W. J., Lopez, I. C., & Soares, J. V. (2012). Leafsnap: A computer vision system for automatic plant species identification. European Conference on Computer Vision, 502–516. Leafsnap Dataset. <http://leafsnap.com/dataset/> (last accessed 2019/04/11).

Lee, S.H., Chan, C.S., Wilkin, P., & Remagnino, P. (2015). Deep-plant: Plant identification with convolutional neural networks. IEEE International Conference on Image Processing, 452–456. doi: 10.1109/ICIP.2015.7350839

Nilsback, M.E., & Zisserman, A. (2008). Automated flower classification over a large number of classes. Sixth Indian Conference on Computer Vision, Graphics & Image Processing, Bhubaneswar, 722-729. doi: 10.1109/ICVGIP.2008.47

Nilsback, M.E., & Zisserman, A. (2006). A visual vocabulary for flower classification. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2, 1447-1454. 17 Category Flower Dataset. <http://www.robots.ox.ac.uk/~vgg/data/flowers/17/> (last accessed 2019/04/11).

Rzanny, M., Seeland, M., Wäldchen, J., & Mäder, P. (2017b). Jena Leaf Images 17, https://doi.org/10.7910/DVN/8BP9L2, Harvard Dataverse, V1. <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/8BP9L2> (last accessed 2019/07/22).

Jadeja, B.A., Odedra, N.K., Solanki, K.M., & Baraiya, N.M. (2006). Indigenous animal healthcare practices in district Porbandar, Gujarat. Indian Journal of Traditional Knowledge, 5, 253-258. http://nopr.niscair.res.in/handle/123456789/6848

Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., et al. (2013). API design for machine learning software: experiences from the scikit-learn project. European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases Workshop: Languages for Data Mining and Machine Learning, Prague, 108–122.