Vol. 5, No. 6 June 2014 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

Flower Species Classification using Statistical and Gist Descriptors

[1] S.Arivazhagan, [2] R.Newlin Shebiah, [3] S.Aarthi, [4] M.Abinaya, [5] S.Sindu Priya
[1 2,3,4,5] Department of Electronics and communication Engineering
Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India

[2] newlinshebiah@yahoo.co.in

ABSTRACT

This Paper aims at classification of flower images by means of Gist descriptor and statistical features using SVM classifier. With advances in digital image processing, automated classification of flower images over large categories of dataset is possible by knowing the name of the flower. In case of unknown species name, classification of flower image can be performed by its visual content. The purpose of the paper is to identify different species of flower with the help of visual content. Here hundred and two categories of flower dataset are used and it is obtained from Visual Geometry Group, university of Oxford. In this paper ten classes are taken from dataset and classified. The Gist descriptor is combined with statistical features such as mean, standard deviation, skewness, Kurtosis are given to SVM classifier. It uses multiple kernel frame work to classify the images. The recognition rate is 79.36%.

Keywords: *Gist descriptor, SVM classifier.*

1. INTRODUCTION

The demand for medicinal plants [1] has increased over the past two decades as herbal medicines are being used widely in traditional medical systems in developed and developing countries. Wrong identification of medicinal plants is one of the factors that make herbal remedy unsafe. Owing to the ignorance of the exact identities of plants used in ayurvedic practice, many exotics are being used mistakenly or as substitutes in the absence of the plants originally recommended. Plants are classified most often on reproductive (flower) characteristics as opposed to vegetative (leaf, stem) characteristics. This has become customary since reproductive parts (petals, sepals, stamens and pistils) remain relatively unchanged over diverse environments, whereas vegetative parts tend to change depending on the environment in which they grow. Identifying a flower using a field guide or key without expert guidance is also a time-consuming task. Furthermore, the fact that some of the flowers being relatively similar and different examples of the same flower differ in colour and shape implies that the recognition by laypersons or pattern recognition systems is not straightforward. Flower recognition becomes cumbersome when using keys as they require answers for a series of questions in order to recognize flowers. These features often relate to the internal structure of the flower which is in most cases visible only when it is dissected. Flower classification [2] is a challenging problem within the field of computer vision that has seen much progress recently. The aim of the paper is to develop a framework that can be applied to the task of flower classification and can be easily extended to other similar classification tasks, i.e. classification of a large number of similar objects which require specialist knowledge. Given an image of a flower, the task is to assign a species label to the image. It is done by analyzing the visual content of the image. Not all species of flowers can be distinguished visually; hence this paper will limit to ones that can be distinguished. The main problem for an object recognition system for any task is to be unaffected by irrelevant variations in the appearance of the object. For object classification, including flower classification, these appearance variations can be due to variations in imaging conditions, object deformations and variations between different object instances of a category.

Saitoh and Kaneko [3] proposed an automatic method for recognizing wild flowers using a frontal flower image and a leaf image taken by a digital camera. A total of 17 features that describe the colour and shape properties of the flower and the leaf images were extracted for flower recognition by using a neural network. A recognition rate of 95% was obtained for the recognition of 20 sets of flower and leaf images from 16species. The main problem with this approach is that it is inconvenient to take the images.

Das et al. [4] proposed an approach to indexing flower patent images using the domain knowledge of flower colours and their spatial locations. The colour features include colour names and their relative proportions in the flower region. Their flower indexing system provided queries using natural language colour names and an example image. A flower database consists of 300 images was tested to demonstrate the effectiveness of their proposed approach.

Nilsback and Zisserman [5] developed a visual vocabulary that explicitly describes the various characteristics (colour, shape, and texture) of flowers. Experimental results on a dataset of 1360 images from 17 flower species have shown that the combined vocabulary outperforms each of the individual ones. Typically, there are too many parameters need to be optimized to get high recognition rate.

2. PROPOSED METHODOLOGY

The proposed methodology has three major categories. Segmentation, Feature extraction and Classification.

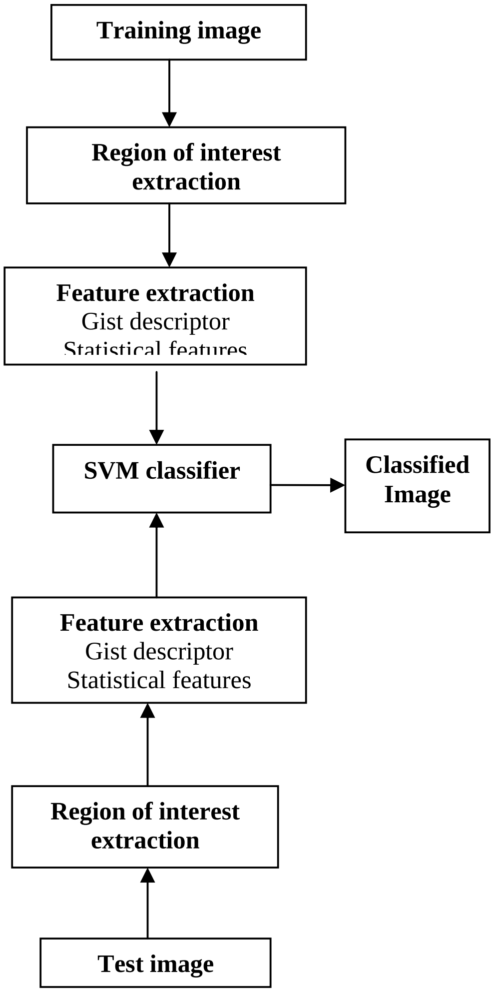

Fig 1: Block Diagram for Proposed Methodology

2.1 Region of Interest Extraction

Thresholding algorithm is used for foreground and background separation. The original image is converted into HSV plane. Threshold is determined by calculating the sum of mean and standard deviation of the intensity plane. Now, the image gets split into foreground and background. Region of interest will be in white colour and background will be in black colour. The original image is decomposed into R, G, B components. Map the threshold image with R, G, B planes separately. Then concatenate the mapped images. The output will be region of interest with black background. Thus background clutter can be removed. Now traverse the pixel values from left side of the image and detect the non zero pixel value. After detecting the location of non zero pixel value, delete the traversed columns up to the detected location. Similarly, traverse it from right, top, and bottom to detect the location and delete the matrix outside the bounding box.

2.2 GIST Descriptor

The inherent problems with large image databases are long access time to read all the image data. Long processing time especially with a per-pixel analysis. To avoid these kinds of problems, descriptors are used. Descriptor is a piece of stored information that is used to identify an item in an information storage and retrieval system.

It save pertinent information, saving the processing time in future queries. It can save image features that are essential for search and comparison. GIST generally means quintessence. It gives essence of an image. It was created for recognition of similar scenes, like mountains, tall buildings, streets.

Fig 2: Inputimage

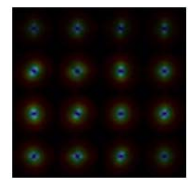

Fig 3: GISTDescriptor

Gist descriptor=n*n*k (1)
Where n*n=no.of.partitions
k=scale* orientation

2.3 Statistical Features

2.3.1 Mean

The mean of a data set is simply the arithmetic average of the values in the set, obtained by summing the values and dividing by the number of values. The mean is a measure of the centre of the distribution.

$$Mean, \mu = \frac{1}{M*N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} p(i,j) \qquad (2)$$

2.3.2 Standard Deviation

The standard deviation is measures of the spread of the distribution about the mean.

$$SD, \sigma = \sqrt{\left(\frac{1}{M*N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left(p(i,j)-mean\right)^2\right)} \qquad (3)$$

2.3.3 Skewness

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the centre point. Measure of Dispersion tells us about the variation of the data set. Skewness tells us about the direction of variation of the data set.

$$S = \frac{E(x-\mu)^3}{\sigma^3} \qquad (4)$$

2.3.4 Kurtosis

Kurtosis is a parameter that describes the shape of a random variable's probability distribution. Kurtosis characterizes the relative peakedness or flatness of a distribution compared to the normal distribution. Positive kurtosis indicates a relatively peaked distribution. Negative kurtosis indicates a relatively flat distribution.

$$K = \frac{E(x-\mu)^4}{\sigma^4} \tag{5}$$

2.4 SVM Classifier

Multi class SVM aims to assign labels to instances by using support vector machines, where the labels are drawn from a finite set of several elements. The dominant approach for doing so is to reduce the single multi class problem into multiple binary classification problems. Common method for such reduction include: Building binary classifiers which distinguish between (i) one of the labels and the rest (one-versus-all) or (ii) between every pair of classes (one-versus-one). Classification of new instances for the one-versus-all case is done by a winner- takes-all strategy, in which the classifier with the highest output function assigns the class (it is important that the output functions be calibrated to produce comparable scores). For the one-versus-one approach, classification is done by a max-wins voting strategy, in which every classifier assigns the instance to one of the two classes, then the vote for the assigned class is increased by one vote, and finally the class with the most votes determines the instance classification. LS-SVM classifiers (Suykens, 1999): close to Vapnik's SVM formulation but solves linear system and QP problem. SVM classifier has decrease rate of convex cost function. The Weighted Version with modified cost function was high in least square SVM classifier. It has high robustness. The original image from Oxford is resized as 256 by 256. Thirteen features are extracted from segmented image and the model is trained. Equal level of images are trained and tested for ten classes.

3. RESULT AND DISSCUSSION

The main problem for an object recognition system for any task is to be unaffected by irrelevant variations in the appearance of the object. These appearance variations can be due to variations in imaging conditions, object deformation etc. It can be eliminated by extracting region of interest. Then features are extracted from the region of interest which will improve classification performance.

Fig 4: Rose

Fig 5: wall flower

Fig 6: Osteospermum

Fig 7: Petuina

Fig 8: Fragipani

Fig 9: Globe thistle

Fig 10: Waterlilly

Fig 11: Passion flower

Fig 12: Bishop

Fig 13: Wild pansy

3.1 Region of Interest Extraction

Thresholding algorithm is used for foreground and background separation. The Original image is converted into HSV plane. Thresholding is determined by calculating the sum of mean and standard deviation of the intensity plane. Now, the image gets split into foreground and background.

Fig 14: Input image

Fig 15: Intensity plane

Fig 16: threshold image

Fig 17: Cropped

Table 1: Recognition of Flower Species Using Statistical Features

SPECIES	Number of Images		Recognition rate in percentage
	Train	Test	
Rose	57	57	50.87
Wallflower	71	71	81.69
Bishop	32	32	68.75
Osteospermum	21	21	47.61
Daisy	44	44	25
Globethistle	25	25	24
Waterlilly	75	75	41
Frangipani	59	59	52.54
Petuina	54	54	38.88
Passion	82	82	70.73
Average			50.14

The above table shows that statistical features provide good classification. Only with statistical features, recognition rate is 50.14% for ten classes.

Table 2: Recognition of Flower Species Using Gist Descriptors

SPECIES	Number of images		Recognition rate in percentage
	Train	Test	
Rose	57	57	54.38
Wallflower	71	71	88.73
Bishop	32	32	65.62
Osteospermum	21	21	57.14
Daisy	44	44	70.45
Globethistle	25	25	84
Waterlilly	75	75	80
Frangipani	59	59	69.49
Petunia	54	54	68.51
Passion	82	82	90.24
Average			72.85

It produces higher classification result when compared with statistical features.

Table 3: Recognition of Flower Species Using Combination of Gist and Statistical Features

SPECIES	Number of images		Recognition rate in percentage
	Train	Test	
Rose	57	57	63.15
Wallflower	71	71	92.95
Bishop	32	32	81.25
Osteospermum	21	21	66.66
Daisy	44	44	81.81
Globethistle	25	25	84
Waterlilly	75	75	84
Frangipani	59	59	71.18
Petuina	54	54	79.62
Passion	82	82	89.02
Average			79.36

When gist descriptor and statistical features are combined, the recognition rate is improved at higher rate.

4. CONCLUSION

This paper discussed a method for classification of flower species based on gist descriptors and statistical features. The flower dataset is taken from visual geometry group, University of Oxford. Region of interest is extracted by means of Thresholding in order to remove the background clutter. Statistical Features such as mean, standard deviation, skewness and kurtosis are extracted. The experimental results show that classification cannot be done with the help of Statistical features alone. But other feature like gist descriptors is needed to improve the recognition rate. This method has recognition rate of 79.36%.

REFERENCES

[1] Ranga Rodrigo, Kalani Samara wickrame and Sheron Mindya," An Intelligent Flower Analyzing System for Medicinal Plants .

[2] M.-E. Nilsback and A. Zisserman."A visual vocabulary for flower classification". In Proc. CVPR, volume 2, pages 1447–1454, 2006.

[3] T Saitoh and T Kaneko,"Automatic recognition of wild flowers ," in15th International Conference on Pattern Recognition, 2000, pp. 507 -510 vol.2.

[4] M. Das, R. Manmatha, and E. M. Riseman. Indexing flower patent images using domain knowledge. IEEE Intelligent Systems, 14(5):24–33, 1999.

[5] Ivan Sikiri´c and Karla Brki´c, SinišaŠegvi´c, "Classifying traffic scenes using gist image descriptors " in Proceedings of the Croatian Computer Vision Workshop, Year 1, September 19, 2013, Zagreb, Croatia.

[6] Varma M, Garg R., Locally Invariant Fractal Features for Statistical Texture Classification. Conf on Computer Vision, 2007, pp. 1-8.

[7] Christopher J.C. Burges,"A tutorial on Support vector machines for Pattern recognition"Kluwer Academic Publishers, Boston

[8] Homepages.inf.ed.ac.uk/rbf/HIPR2/classify.htm.

[9] Wacker, A.G and Landgrade, D.A., "Minimum distance classification in Remote sensing", LARS Technical Reports, Paper 25.

Vol. 5, No. 6 June 2014 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

[10] en.wikipedia.org/ wiki/support vector Machine.

[11] M.-E.Nilsbackand A. Zisserman, "An Automatic visual Flora segmentation and classification of flower images".

[12] J. Ye and T. Xiong, SVM versus least squares SVM, in Proceedings of the International Conference on Artificial Intelligence and Statistics. Citeseer, pp. 640–647, 2007.

[13] T. Saitoh, K. Aoki, and T. Kaneko.Automatic recognition of blooming flowers. In Proc. ICPR, volume 1, pages 27–30,2004.

[14] M.-E. Nilsback and A. Zisserman."Delving into the whorl of flower segmentation". In Proc. BMVC., volume 1, pages 570–579, 2007.

[15] Yuning Chai," Recognition between a Large Number of Flower Species",Master Thesis Swiss Federal institute of Technology, Zurich.

[16] Maria-Elena Nilsback and Andrew Zisserman,"Automated flower classification over a large number of classes, Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing,2008.

[17] Johan Suykens,"Support vector machines and kernel methods insystems, modelling and control",http://www.esat.kuleuven.be/scd/

[18] G.R.G.Lanckriet,L.E.Ghaoui,C. Bhattacharyya, and M.I. Jordan. A robust minimax approach to classification. Journal of Machine Learning Research, pages 555-582, 2002.