Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 58 (2015) 723 – 730

Procedia Computer Science

Second International Symposium on Computer Vision and the Internet (VisionNet'15)

Automatic Identification of Licorice and Rhubarb by Microscopic Image Processing

Bhupendra Fataniya[a], Meet Joshi[b], Urmil Modi[c], Tanish Zaveri[d]

[a,b,c,d] *Nirma University, Ahmedabad, 382481, India*

Abstract

This paper presents a method for automatic identification of Herbal Plants Licorice and Rhubarb by microscopic image processing. This method is useful for identifying species from fragments or powders and for distinguishing species with similar morphological characteristics. In first step, desired region of the image is cropped by intensity based segmentation. After that Hu's moments and compactness parameters are calculated for cropped part which are rotational, scaling and shift invariant. In the final stage using SVM classifier, classify the herbal plants of licorice and rhubarb. Area under the ROC curve for licorice is 0.7051 and for rhubarb it is 0.9487.

Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)

Keywords:M oments; Segmentation; compactness parameter; SVM classifier; ROC curve

Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)
doi:10.1016/j.procs.2015.08.093

1. Introduction

Microscope image processing is a broad term that covers the use of digital image processing techniques to process, analyze and present images obtained from a microscope[1]. Image processing techniques have been widely used in the last decade in medical imaging and the microscopic field. Herbal drug technology includes all the necessary steps for the conversion of botanical materials into medicines, where standardization and quality control are essential analytical tools to assure the correct identification of drugs. Advances in microscope technology and improvements in light and scanning electron microscopes have increased the accuracy and capabilities of microscopy as a mean of botanical identification[2]. Microscopy allows the identification of herbal drugs and the detection of individual components of a mixture[2]. It is highly important to ensure quality and purity of herbal medicines in order to maximize the efficacy and minimize adverse side effects[2]. For these reasons, microscopic analysis of the powdered herbal drugs is a mandatory technique on the official Pharmacopeia monographs.

To this day, analysis of microscopic images of human tissues remains the most reliable way of diagnosing and grading several illnesses, e.g., cancer. Our work on the classification and rating of microscopic images is motivated by the fact that there is significant inter- and intra-rater variability in the grading and diagnosis of these illnesses from image data by human experts. Computer-aided classification can assist the experts in making their grading and diagnosis reproducible, while additionally providing useful quantitative measures.

One of the first applications of computer-aided diagnosis (CAD) was digital mammography[3, 4] in the early 1990's . CAD has become a part of routine clinical detection of breast cancer on mammograms at many screening sites and hospitals[5] in the United States. Researchers in both the image analysis and pathology fields have recognized the importance of quantitative analysis of pathology images. Since most current pathology diagnosis is based on the subjective (but educated) opinion of pathologists, there is clearly a need for quantitative image based assessment of digital pathology slides[6].

In the present scenario, the work of identifying the plant in its powdered form is being done by pharmaceutical companies who perform this task manually[7]. The experts analyze the powder and decide which plant it is[8]. This process consumes a more time and it is also less efficient. In this paper, we propose a system to analyze the microscopic images of given plants licorice and rhubarb which is in powder form and identify the plant based on its characteristic features automatically.

The compound glycyrrhizic acid, found in Licorice, is used for treatment and control of chronic viral hepatitis[9]. Licorice affects the body‘s endocrine system. It is also a mild laxative and may be used as a topical antiviral agent for shingles, ophthalmic, oral or genital herpes.

Rhubarb stems are used as food. Rhubarb can be used as a strong laxative. The rhizomes contain stilbene compounds which seem to lower blood glucose levels in diabetic mice[10].

The whole algorithm is divided into two steps. The first step involves conversion of colour image into greyscale image, contrast adjustment, image thresholding, morphological operations and object cropping. At the end of first image we get automatically segmented object.

Second step involves feature extraction and classification. In feature extraction Hu moments[11] and compactness parameters[12] are calculated for the cropped object. So, this makes the algorithm scale, shift and rotation invariant[8]. For the classification Support Vector Machine is used.

The rest of the paper is organized as follows: Section 2 discusses the proposed algorithm. Followed by Section 3 that discusses the segmentation method and challenges faced in automatic segmentation. Section 4 discusses the feature extraction which includes Hu's moments and compactness parameter. Section 5 is about the SVM classifier and roc curve. Section 6 experiments and their results and section 7 concludes the paper.

2. Proposed Algorithm

In this paper our aim is to provide new framework for microscopic image analysis and intelligent identification systems based on its optimal extracted feature set. As described above we have considered two herbal plants to apply our proposed algorithm. The proposed algorithm is shown in Fig. 1.

The characteristic feature of Licorice is the shape of its crystal phloem fiber and that of rhubarb is its crystal structure as shown in Fig. 2(a). We had 78 and 52 microscopic images of licorice and rhubarb respectively. The algorithm is described in following six steps.

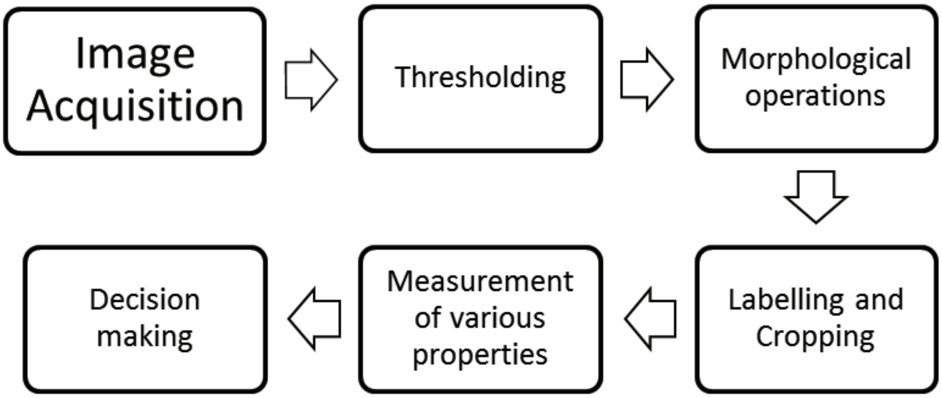


Fig. 1. Block diagram of proposed algorithm

- Image Acquisition: 78 licorice images and 52 rhubarb images are obtained using the microscope and by application of operations like scaling, rotation and both to the images.
- Thresholding: These images are enhanced in spatial domain by conversion to gray scale followed by application of BCV (between class variance) threshold.
- Morphological Operations: By setting a constant threshold, morphological operations are applied on the image to remove pixels or pixel areas smaller than a predefined value.
- Labelling and cropping: Objects in the image so obtained are labelled and separate images are formed for each object.
- Measurement of various properties: Various physical characteristics of the feature are extracted and measured. Moments and compactness factor are used as a feature for the classification of microscopic images.
- Decision Making: The data so obtained is fed to algorithm of SVM classifier thus classifying the test image into either rhubarb or licorice.

3. Automatic Segmentation

In the first step the colour image is converted into a greyscale image. Single pixel of this greyscale image contains an 8 bit number which indicates the intensity of the pixel between 0 to 255 levels. After that contrast adjustment is applied so the black pixel becomes more black and white pixel becomes whiter. After that by using "Otsu" thresholding[10] technique the greyscale image is converted into a binary image. Single pixel of this binary image contains a 1 bit number which indicates the intensity of the pixel either 0 or 255. Threshold Intensity value is 168 for licorice and 145 for rhubarb. By using morphological operations such as opening, closing, dilation, erosion etc noise is removed. After that the desired object is cropped. Results of the automatic segmentation of Licorice and rhubarb are shown below in Fig. 2.

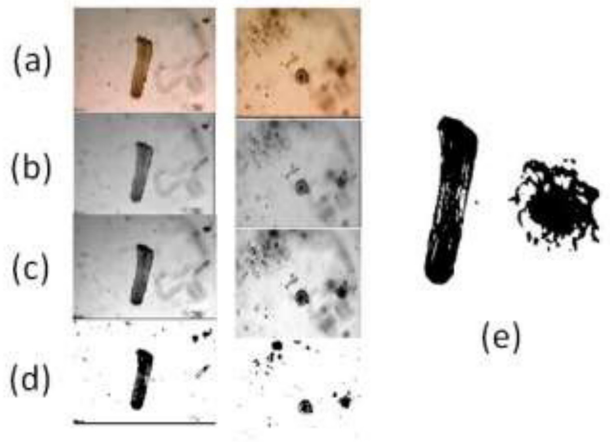


Fig. 2. Automatic Segmentation Result : (a) Original Image, (b) Converted into Grayscale, (c) Obtained after contrast Adjustment, (d) Obtained after OTSU Thresholding (d) Cropped Object

Sometimes when the image is blurred or the cell is broken or the cell is hidden under other objects we get some error in the segmentation. Examples of such licorice and rhubarb images are given below Fig. 3.

Fig. 3. Challenges in segmentation

4. Feature Extraction

The first step of most image interpretation problems including microscopic image processing is feature extraction. Feature vectors represent or summarize a given image. In many problems, image classification is then performed using the feature vectors instead of actual pixel values. Microscopic images are large in size, which requires the feature extraction process to be computationally efficient in order not to rely on supercomputers or computer clusters in the future.

Feature calculated for the segmented object should be robust, discriminate, reliable, independent, rotation invariant, scale invariant and scale invariant. Moments[13] are scalar quantities used to characterize an image and to capture its significant features. Image moments are useful to describe objects after segmentation. We have considered two features to represent the image.

4.1 Moments[13, 15]:

Hu's Moments are basically a set of features for image description which are invariant to rotation, shifting and scaling operations.
General moment of an M*N image f(x , y) is

$$m_{pq} = \Sigma\,\Sigma\, x^p y^q f(x , y) \quad (1)$$

where m_{pq} is moment of order (p+q).
Centralized moments of f(x , y) is

$$\mu_{pq} = \Sigma\,\Sigma\, (x-\overline{x})^p(y-\overline{y})^q f(x , y) \quad (2)$$

where $\overline{x} = m_{10}/m_{00}$ and $\overline{y} = m_{01}/m_{00}$
Normalized central moment is

$$n_{pq} = \mu_{pq} / \mu_{00}^{\gamma} \quad (3)$$

where $\gamma = (p + q + 2)/2$

Seven Hu's moments[13] are (4)

$$M1 = \eta_{20} + \eta_{02}$$
$$M2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$
$$M3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$
$$M4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$
$$M5 = (\eta_{30} - 3\eta_{12}) (\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 -3 (\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03}) (\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 -(\eta_{21} + \eta_{03})^2]$$
$$M6 = (\eta_{20} - \eta_{02}) [(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12}) (\eta_{21} + \eta_{03})$$
$$M7 =(3\eta_{21} - \eta_{03}) (\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 -3 (\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03}) (\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 -(\eta_{21} + \eta_{03})^2]$$

4.2 Compactness Parameter[12]:

Compactness, the first shape factor, is a simple measure of the efficiency of a contour to contain a given area, defined by

$$C_0 = P^2/A \quad (5)$$

Where *P* is perimeter of the object and *A* is area enclosed by that. The value of compactness is larger when the area contained by contour of given length is smaller. The definition of compactness is modified to restrict the range of parameter to interval [0,1] as

$$cf = 1 - 4\pi A/P^2 \quad (6)$$

For 10 licorice and 10 rhubarb images M1, M2 and M5 hu moments and cf values are shown in below Table 1 and 2. From this we can say that moments M1, M2 and M5 and cf are invariants in the same class. Here we have consider only three moments in which variations is less.

Table 1 Hu moments and cf for Licorice

No.	M1	M2	M5	Cf
1	2.960	4.38	63	0.590
2	2.09	4.38	62.6	0.623
3	2.09	4.38	63	0.590
4	2.10	4.42	64.5	0.580
5	0.873	0.763	0.244	0.499
6	0.866	0.750	0.232	0.587
7	0.855	0.731	0.215	0.496
8	2.09	4.38	62.9	0.588
9	2.09	4.38	62.9	0.588
10	2.08	4.35	61.3	0.590

Table 2 Hu moments and cf for Rhubarb

No.	M1	M2	M5	Cf
1	1.37	1.87	3.72	0.927
2	1.37	1.88	3.78	0.944
3	1.37	1.89	3.83	0.952
4	1.39	1.93	4.14	0.924
5	1.34	1.81	3.37	0.927
6	1.32	1.73	2.97	0.924
7	1.26	1.59	2.30	0.921
8	1.35	1.82	3.47	0.919
9	1.33	1.77	3.16	0.906
10	1.28	1.64	2.54	0.890

5. SVM classifier[14,16]:

Classifying data has been one of the major parts in machine learning. The idea of support vector machine is to create a hyper plane in between data sets to indicate which class it belongs to. The challenge is to train the machine to understand structure from data and mapping with the right class label, for the best result, the hyper plane has the largest distance to the nearest training data points of any class.

We have used the moments and compactness factor as a feature for the object recognition and classify the sample using SVM classifier.

Receiver operating characteristics (roc curve) is a graphical representation illustrating the performance of a classifier. This curve is traced by plotting true positive rate against false positive rate at various threshold ratings. True positive rate is also known as sensitivity and false positive rate is also known as specificity.

$$\text{Sensitivity} = TP / (TP + FN) \quad (7)$$

$$\text{Specificity} = TN / (FP + TN) \quad (8)$$

6. Experiments and Results

For the experiment, we have 78 licorice and 52 rhubarb powder microscopy images. In the data base we have included rotation, scaling and shifted powder microscopic images. Some of the images from the data base used as training image and after that we apply all images for the testing. We used the hu's moments (M1, M2 and M5) and Cf values for the features and apply two class SVM classifier for the classification. For the validation of classification, we have prepared the confusion matrix and ROC curve. Algorithm implemented in MATLAB environment and system is Intel i3 CPU, 2.4 GHz, and 64 bit 4 GB RAM.

Images were divided into training and testing sets. For training 28 licorice images and 28 rhubarb images and for testing 78 licorice images and 52 rhubarb images were taken.

In the first experiment Hu moments M1 and M2 were taken as the parameters and in the second experiment Hu moment M5 and compactness parameter cf were taken as parameters. Confusion Matrices are given below for both the experiments in Table 3 and 4.

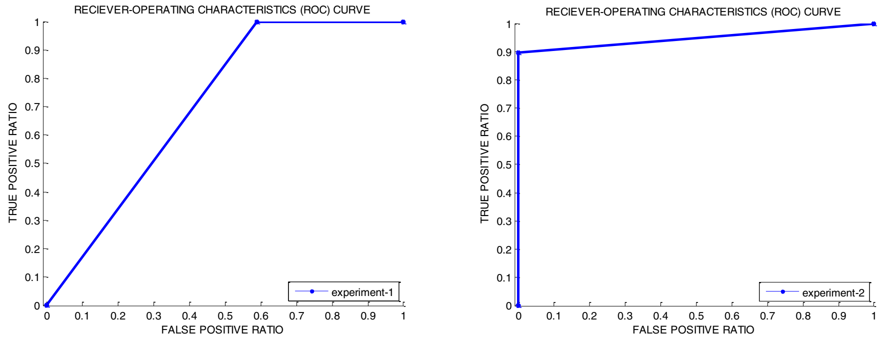
Roc curve for classification of licorice and rhubarb is shown below in Fig. 4. Area under the ROC is 0.7051 for licorice and for rhubarb it is 0.9847.

Table 3 Confusion Matrix For Rhubarb (Parameters: M1 & M2)

SVM	Positive	Negative	Total
True	52	0	52
False	46	32	78

Table 4 Confusion Matrix For licorice (Parameters: M5 & cf)

SVM	Positive	Negative	Total
True	70	8	78
False	0	52	52



(a) (b)

Fig. 4. ROC curve for (a) licorice and (b) rhubarb

7. Conclusion

Features based image identification is an effective way of detecting and analyzing microscopic images of herbal plant. It can save a lot of time and effort. In this paper we have implement the automatic segmentation using thresholding and finding moments and compactness factor as a features which is rotational, scaling and shifted invariants. Finally using SVM classifier, we classify the licorice and rhubarb and area under the ROC is 0.7051 and

0.9875 achieved for licorice and rhubarb. We can improve ROC by including all seven moments and other invariants features.

Acknowledgements

The author acknowledge the GUCOST for the funding this project and management of Nirma university for allowing to present this paper.

References

1. M. D'Acunto, O. Salvetti. Image analysis aided by Segmentation algorithms for techniques ranging from Scanning Probe Microscopy to Optical Microscopy. National Research Council, Istitute of Information Science and Technologies, Pisa,Italy.
2. R. Serrano, G. da Silva and O. Silva. Application of Light and Scanning Electron Microscopy in the Identification of Herbal edicines, Microscopy: Science, Technology, Applications and Education, 2010
3. A. Meacutendez, P. Tahoces, M. Lado, M. Souto, and J. Vidal, "Computeraided diagnosis: automatic detection of malignant masses in digitized mammograms," *Med Phys*, vol. 25, no. 6, pp. 957–64, 1998.
4. M. N. Gurcan, Y. Yardimci, A. E. Cetin, and R. Ansari, "Automated detection and enhancement of microcalcifications in mammograms using nonlinear subband decomposition," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'97),*, vol. 4, pp. 3069 – 3072 vol.4, apr 1997.
5. J. Tang, R. Rangayyan, J. Xu, I. El Naqa, and Y. Yang, "Computeraided detection and diagnosis of breast cancer with mammography: recent advances," *Trans. Info. Tech. Biomed.*, vol. 13, no. 2, pp. 236–251, 2009.
6. M. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147 –171, 2009.
7. Bhupendra Fataniya, Prachi Patel, Tanish Zaveri and Sanjeev Acharya. Microscopic Image analysis method for identificxation of indian herbal plants, International conference on Devices, Circuits and Communications(ICDCCom-2014).
8. World Health Organization. Good Manufacturing Practices for pharmaceutical products: main principles. WHO Technical Report Series, No. 908, 2003
9. International food quality research institute. Licorice industry in china : Implication for licorice producers in uzbekistan.
10. Radhika R., Dr,Krishnakumari, Dr.Sudarsanam D. Antidiabetic activity of Rheum emodi in Alloxan induced diabetic rats.
11. Ming-Kuei Hu. Visual Pattern recognition by moment invariants. IEEE xplore.
12. Dijana Tralic, Jelena Bozek, Sonja Grgic. Shape Analysis and Classification of Masses in Mammographic Images using Neural Networks
13. Zhihu Huang, Jinsong Lang. Analysis of Hu 's Moment Invariants on Image Scaling and Rotation.
14. Durgesh K. Srivastava, Lekha Bhambhu . Data classification using support vector machine.
15. H. Ming-Kuei, "Visual pattern recognition by moment invariants," *Information Theory, IRE Transactions,* vol. 8, pp. 179-187, 1962.
16. Cortes, V. Vapnik, "Support vector networks". Journal of Machine Learning, 20, 1995.