Chromatogram Image Pre-Processing and Feature Extraction for Automatic Soil Analysis

Saritha.V, Minu Mary Joseph, Sukhendu Das, Deepak Khemani
Department of Computer Science and Engineering,
Indian Institute of Technology Madras, India.
{sarithav, minumj}@cse.iitm.ernet.in, {sdas, khemani}@iitm.ac.in

Abstract

A circular paper chromatogram is obtained from an alkaline solution of silver nitrate and soil. The shape, size, color and textural patterns of the chromatogram image are hypothesized to contain important information of the mineral content in the soil. We present a method to automatically analyze the chromatogram image for feature extraction. Image pre-processing is an important step before extracting the features of the image. Chromatogram image preprocessing involves detecting the center of the chromatogram, normalization and then segmentation into different concentric regions. Since chromatogram patterns are similar to iris (human eye) patterns, we have adopted iris-preprocessing methods. In this paper, we present a combination of different approaches: to detect the center, normalize and segment the chromatogram. Centre detection algorithm finds the center of the chromatogram which is assumed as the origin for normalization. Chromatogram normalization involves transforming from Cartesian to polar coordinates, so that chromatogram looks like an unwrapped polar image. Finally, color texture segmentation is used to detect different regions. Results of feature extraction are compared to that given by soil experts to test the accuracy of the system.

1. Introduction

A circular paper chromatogram is obtained from an alkaline solution of silver nitrate and soil. The shape, size, color and textural patterns of the chromatogram image are hypothesized to contain important information of the mineral content in the soil. A sample chromatogram is shown in figure 1. Most chromatograms contain two (or three) concentric regions, inner and middle (although some may have a outer region). The method proposed in this paper will be new, as the current process of chemical analysis of soil is done manually, which is an expensive, time consuming and laborious process. This method of analysis will benefit farmers all across the globe and more so in India, who are looking for innovative means to obtain their soil characteristics during the process of farming. This chromatogram based approach would be simpler and much cheaper as compared to manual analysis.

Figure 1. A sample chromatogram image

The focus of this paper is on chromatogram image preprocessing, which is a fundamental step for feature extraction. Since the chromatogram pattern looks like an iris pattern, we adopted iris preprocessing techniques for the purpose of chromatogram preprocessing. John Daughman developed many algorithms for iris recognition [1]. Later many researchers started working in iris image preprocessing which is an important step in iris recognition. Iris image preprocessing includes detection of the pupil center [8], normalization [1] & [5] and segmentation. To detect chromatogram center we used the method [8] of using projection to locate the center. Papers for texture segmentation [3], [4] & [7] use a combination of Gabor and Discrete Wavelet Transform (DWT) based features to segment different texture regions in synthetic and aerial real-world images.

2. Chromatogram Preprocessing

The different regions in a chromatogram image appear as concentric circular regions, which are however not perfectly circular. Since the geometrical features of the regions are vital for the analysis of mineral content in soil, the unwrapping of the original image helps in forming rectangular regions for which the features can be computed easily. Also the problem of identification of different circular regions is simplified to the detection of different rectangular regions arranged in a vertical stack. A normalization process is implemented by preprocessing to facilitate the segmentation and feature extraction, by transforming image from cartesian to polar co-ordinate. This transformation process will produce chromatogram regions, which have the same constant dimensions so that characteristic features will appear to be at the same spatial location and circular regions of the chromatogram appears as near-rectangular (horizontal) segments which is shown later in this paper. The original image is transformed into a polar image using the center of the chromatogram as origin.

2.1. Center Detection

Every chromatogram contains a hole which can be identified against a distinct background. We use a black background to locate the hole. Since the center of the chromatogram is located within the hole and gray information of the hole can be used to detect the center, we decided to consider only that region which covers the hole for center detection. We crop a portion of the image with a fixed window size so that it covers the hole of a chromatogram.

As shown in figure 2(a) the intensity of the hole is lower than the surrounding regions of chromatogram image map. A fixed threshold of about 90 can isolate the hole from the other part. With this threshold a binary image can be achieved by applying equation (1), which is shown in figure 2(b).

$$B(x,y) = \begin{cases} 0 & \text{if } I(x,y) > \alpha \\ 1 & \text{if } I(x,y) \le \alpha \end{cases} \qquad (1)$$

where $I(x,y)$ = Intensity value at pixel (x,y) and α is a threshold value

In the binary image using equation (2), we can locate the center of the chromatogram approximately. Detected center is shown in the figure 2(c) marked by a red circle.

$$x_c = x_i \ \text{ if } \sum_{y=1}^{n} B(x_i, y) = \max\Big(\bigcup_{x=1}^{m} \sum_{y=1}^{n} B(x, y) \Big)$$

$$y_c = y_i \ \text{ if } \sum_{x=1}^{m} B(x, y_i) = \max\Big(\bigcup_{y=1}^{n} \sum_{x=1}^{m} B(x, y) \Big) \qquad (2)$$

where m, n =size of window

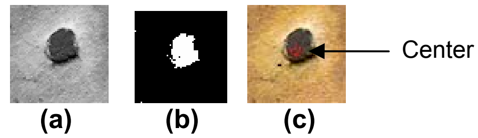


Figure 2(a) Image map of the center of the chromatogram in figure1. (b) Binary image map of (a), after thresholding. (c) Detected center of (a)

2.2. Transform to Polar Image

Once the center of the chromatogram is found, the original image is transformed to polar co-ordinate using this center as origin. This transformation is done using following equation (3).

$$I(x(r,\theta), y(r,\theta)) \rightarrow I(r,\theta) \qquad (3)$$

where
$$x(r,\theta) = r * \cos\theta + x_c$$
$$y(r,\theta) = r * \sin\theta + y_c$$

where (x_c, y_c) is the co-ordinate of the center of the chromatogram, M is maximum radius which depends on the size of the image and $0 \le \theta \le 2\pi$, and $0 \le r \le M$. Figure 4 shows the polar image of chromatogram in figure 1.

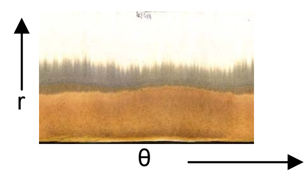


Figure 4. Polar image of the chromatogram in figure 1.

3. Chromatogram Segmentation and Feature Extraction

Color texture segmentation is performed to detect different regions of the chromatogram. The steps of the overall methodology for texture classification [6] are shown in figure 5. The image is filtered using dyadic discrete wavelet transforms [3]. The filter coefficients (responses) are post-processed using a set of non-linear functions, which compute the local energy estimates of the filtered coefficients. These non-linear functions consist of two stages: (i) obtaining the absolute magnitude followed by (ii) smoothing by a large

Gaussian function. These steps are described in the section 3.1. The feature vectors computed from the local energy measure estimates are local mean and local variance, which represent local texture characteristics. These feature vectors are computed from the various filtered images and provided to the Fuzzy C-means (FCM) algorithm [2] to segment the texture patterns in the image. In our experiment, user provides the desired number of classes as an input to the classifier.

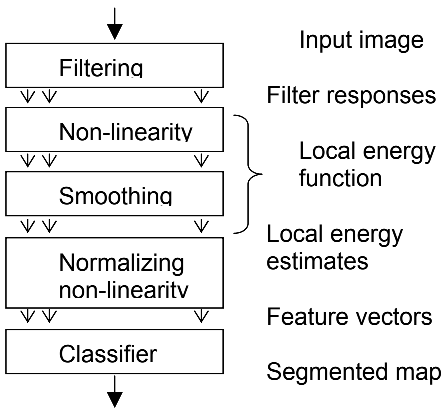


Figure 5. Stages of processing for texture classification

3.1 DWT Features for Classification

The DWT analyses a signal based on its content in different frequency ranges. Therefore it is very useful in analyzing repetitive patterns such as texture.

The 2-D transform uses a family of wavelet functions and its associated scaling function to decompose the original image into different channels, namely the *low-low*, *low-high*, *high-low* and *high-high* (*A,V,H,D* respectively) channels. The decomposition process can be recursively applied to the low frequency channel (*LL*) to generate decomposition at the next level. Figure 6(a), (b) show the 2-channel level-2 dyadic DWT decomposition of an image. The LP and HP *filters* are used to implement the wavelet transform.

The features are computed as the local energy of the filter responses. A local energy function is computed consisting of a non-linearity, by rectifying the filter response and smoothing. Rectification is understood as the operation of transforming negative amplitudes to the corresponding positive amplitudes. Commonly applied smoothing filters in the local energy function are rectangular and Gaussian. Gaussian filter is the better choice and will consequently be used in our experiments.

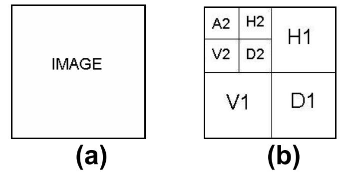


Figure 6. Level 2-dyadic DWT decomposition of an image

The filter coefficients are post-processed using a set of non-linear functions, which compute the local energy estimates (as shown in figure 5). The absolute magnitude of the filter output is taken as:

$$a_j(x,y) = |\, ih_j(x,y) \,|$$

where $ih_j(x,y)$ is the j^{th} channel output of the filter. A low pass Gaussian post-filter $g_p(x, y)$ is then applied to each $a_j(x, y)$ yielding post-filtered energy of the j^{th} filter channel as:

$$e_j(x, y) = a_j(x, y) ** g_p(x, y)$$

where, $$g_p(x,y) = \frac{1}{2\pi\sigma_2^2} e^{\frac{-[x^2+y^2]}{2\sigma_2^2}}$$

where ** denotes convolution in 2-D.

The feature vectors computed from the local energy estimates are (i) mean of A, $\mu[e_j(x, y)]$ and (ii) variance of V and H, $\sigma[e_j(x, y)]$. Hence every pixel in the image is represented by a 9-dimensional (3*3) feature vector (3 color bands, 3 DWT subbands).

3.2 Segmentation using FCM

The output from the local energy function is a set of images, one image per filter. These images are feature images that will form the basis for the classification. They are used to form feature vectors, where each feature image corresponds to one element in the feature vector. Each feature vector corresponds to one or a few image pixels. Classification is the task of assigning class labels to these feature vectors.

There are already a large number of supervised and unsupervised texture segmentation algorithms existing in literature. The difference between supervised and unsupervised segmentation is that supervised segmentation assumes priori knowledge on the type of textures present in the image. We have used here, the fuzzy c-means clustering (FCM) algorithm as an iterative procedure, which is described below:

1. Calculate the fuzzy cluster centers v_c^l with

$$v_c = \sum_{m=1}^{M} (u_{c,m})^W x_m \Big/ \sum_{m=1}^{M} (u_{c,m})^w$$

2. Update $U^{(l)}$ with

$$u_{c,m} = 1 / \sum_{m=1}^{C} \left(\frac{d_{c,m}}{d_{j,m}} \right)^{\frac{2}{W-1}}$$

where $(d_{i,m})^2 = \left\| x_m - v_i \right\|^2$ and

$\left\| \bullet \right\|$ is any inner product induced norm

3. Compare with $U^{(l+1)}$ in a convenient matrix norm.

If $\left\| U^{(l+1)} - U^{(l)} \right\| \leq \varepsilon$ Stop

Otherwise return to Step 1.

M is the size of input data {x_m; m = 1 ,.., M}, C is the number of clusters, w is the fuzzy weighting exponent ($1<w<\infty$) and $U^{(l)}$ *is* the membership function matrix at iteration *l*. The value of the weighting exponent *w* determines the fuzziness of the clustering decision. A smaller value of *w*, i.e. *w* close to unity, will lead to a zero/one hard decision membership function, while a larger *w* corresponds to a fuzzier output.

3.3. Chromatogram Feature Extraction

Most of the chromatograms contain three regions inner, middle and outer as classified by experts. The size, color, shape and textural patterns of these regions are hypothesized to contain important information of the mineral content of soil. We hence extract the features such as width, area, and color of each segmented region.

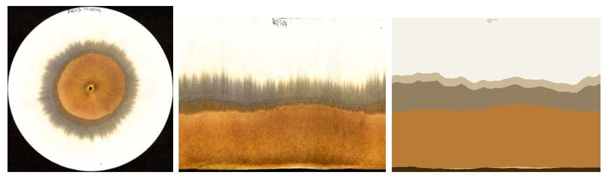
For each region or strip in the segmentation output, we compute the following features:

Area = Number of pixels in that region;
Width = Area / N,
N = Number of columns in the polar image;
Area (cm^2) = Area / Resolution^2 (pixel/cm);
Width (cm) = Width / Resolution (pixel/cm);
Color = Mean of multi-band intensity of the pixels in that region.

4. Experimental Results

In our experimentation we have used 700 chromatograms images to test our algorithm. We have manually grouped the chromatograms based on visual similarity. This helps in comparing results of similar chromatogram producing , identical segmentation results.

The result of segmentation of the polar image of the chromatogram, in figure 4, is shown in figure 7(c). Table 1 shows the features extracted from each segmented strip in figure 7(c). A similar example of this process is shown in figure 8 and table 2. We neglect the narrow and thin strips of segments formed in the output (as in figure 7(c) and 8(c)). These segments are irrelevant in the characteristics of the chromatogram which reflect the soil content. Hence, tables 1 and 2 give the features of the prominent segments only. The advantage of chromatogram preprocessing (conversion from Cartesian to polar) is now evident, as one can visually compare the spread of the significant regions and compute with ease the geometrical features necessary for soil analysis. In the original circular chromatogram image, these computation would not have been trivial.

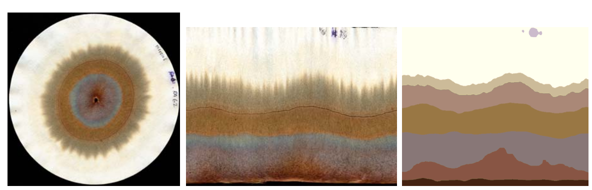


(a) (b) (c)

Figure 7. a) Chromatogram image, b) Polar image of chromatogram in (a), c) Segmentation of polar image (b).

Table 1. Features of the chromatogram segments in figure 7(c).

Segment Label	Area (cm^2)	Width (cm)	Color		
			Red	Green	Blue
Back Ground	23.78	2.938	254	252	244
Middle	8.705	1.076	175	160	137
Inner	23.23	2.875	206	157	102



(a) (b) (c)

Figure 8. a) Chromatogram image, b) Polar image of chromatogram in (a), c) Segmentation of polar image (b).

Table 2. Features of the chromatogram segments in figure 8(c).

Segment Label	Area (cm^2)	Width (cm)	Color		
			Red	Green	Blue
Back Ground	20.37	2.475	250	247	235
Outer	7.19	0.874	182	167	144
Middle	9.99	1.215	176	147	115
Inner	17.83	2.166	161	141	130

4.1. Analysis of Results

We evaluated the success rate for the proposed method using 50 chromatograms with diverse textures, from 500 chromatograms given by experts. For the two chromatograms in figure 7 & 8, comparison of the experimental results with the data given by soil experts is listed in the table 3. The errors are shown separately for the images in figure 7(a) (two regions) and figure 8(a) (three regions). Our experimental results were compared with those obtained manually by experts and the mean and maximum errors obtained for 50 chromatograms are shown in table 4. Mean error is about 1-2mm while the maximum is about 5mm compared to the chromatogram size of a radius of 75mm (approximately).

Table 3. Comparisons of experimental result with expert data.

Region	Width (in cm)		Error (in cm)
	By experts	By proposed method	
For Chromatogram in figure 7			
Inner	2.86	2.8716	0.0116
Middle	1.25	1.076	0.174
For Chromatogram in figure 8.			
Inner	2.186	2.166	0.02
Middle	1.3	1.214	0.086
Outer	0.9	0.874	0.026

Table 4. Error (in cm) analysis for 50 chromatograms

Region	Mean error	Maximum Error
Inner	0.227	0.378
Middle	0.131	0.532
Outer	0.158	0.321

5. Conclusion

We have presented a method to automatically analyze the chromatogram before extracting the features of the image. Chromatogram image preprocessing involves combination of different modified approaches, detecting the center of the chromatogram, normalization and then segmentation into different circular regions followed by extraction of features. Experts have validated the results of feature extraction and they are convinced that these features will help to automate the analysis of the mineral content of the soil. As a future scope of work, a knowledge base using concepts of case-based reasoning is being built from the features and information extracted out of the chromatogram image analysis. This knowledge base would help us to analyze the mineral content of different types of soils, and thus develop an automated system for soil analysis.

6. Acknowledgement

This work is sponsored by DST (GOI), New Delhi. Chromatogram samples are provided by Murugappa Chettiar Research Center, Chennai-600036.

7. References

[1] John Daugman, “High Confidence Visual Recognition of Persons by a Test of Statistical Independence”, In *Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1993, pp. 1148-1161.

[2] M. F. A. Fauzi and P. H. Lewis, “A Fully Unsupervised Texture Segmentation Algorithm”, In *Proceedings of British Machine Vision Conference*, Norwich, UK, 2003, pp. 519-528.

[3] C. Lu, P. Chung and C. Chen, “ Unsupervised Texture Segmentation Via Wavelet Transform”, *Pattern Recognition*, 1997, pp. 729-742.

[4] Li Ma, Tieniu Tan, Yunhong Wang and Dexin Zhang, “Personal Identification Based on Iris Texture Analysis”, In *Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence*, December 2003, pp. 1519-1533.

[5] Ahamad Poursaberi and Babak N. Araabi, “A Novel Iris Recognition System Using Morphological Edge Detector and Wavelet Phase Features”, In *Proceedings of the ICGST International Journal on Graphics,Vision and Image Processing,* Cairo, Egypt, December 2005, pp. 9-15.

[6] Trygve Randen and John Hakon Husoy, “Filtering for Texture Classification, A Comparative Study”, In *Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence*, April 1999, pp. 291-310.

[7] Shivani G. Rao, Manika Puri, and Sukhendu Das, “Unsupervised segmentation of Texture Images Using a Combination of Gabor and Wavelet Feature”, In *Indian Conference on Computer Vision, Graphics and Image Processing,* Kolkata, India, December 2004, pp. 370-375.

[8] Qi-Chuan Tian, Quan Pan, Yong-Mei Cheng, and Quan-Xue Gao, “Fast Algorithm and Application of Hough Transform in Iris Segmentation”, In *Proceedings of Third International Conference on Machine Learning and Cybernetics*, Shanghai, 26-29 August 2004, pp. 3977-3980.