Automatic Lane Detection in Chromatography Images

Bruno M. Moreira[1]
moreira@fe.up.pt

António V. Sousa[1,2]
ats@isep.ipp.pt

Ana M. Mendonça[1,3]
amendon@fe.up.pt

Aurélio C. Campilho[1,3]
campilho@fe.up.pt

[1] Laboratório de Sinal e Imagem, Instituto de Engenharia Biomédica, Universidade do Porto.

[2] Instituto Superior de Engenharia, Instituto Politécnico do Porto

[3] Faculdade de Engenharia, Universidade do Porto

Abstract

This paper describes a method for automating the detection of lanes in chromatography images. This is a relevant component of a screening tool for Fabry disease, which will be based on the automatic analysis of the chromatographic patterns extracted from each one of the detected lanes. Our approach includes a pre-processing step resulting in a smoothed profile that is the input of a lane detection step. The proposed method was tested using 66 chromatography images with very promising results.

1 Introduction

Fabry disease (FD) is a Lysossomal Storage Disorder originated from a deficiency in α-galactosidase A, leading to an abnormal accumulation of glycosphingolipids, namely Gb3 [1]. The complete diagnosis of FD is very complex but the first phase is simply based on the detection of an abnormal quantity of Gb3 in urine or blood plasma of the patient. The direct measurement of those compounds can be carried out by using micro tandem mass spectrometer (MS/MS), but their use is very expensive. Another approach, less expensive, is the analysis of a patient urine sample or blood plasma, performed by a Thin-Layer Chromatography (TLC) on a silica gel plate, followed by a visual inspection of the generated chromatographic pattern [2].

In order to implement a screening tool for FD, we need to develop several procedures for automating the complete image analysis process. One fundamental initial step is the detection of the lanes associated with individual samples. This is a fundamental phase as the lanes in the chromatography images contain the composition and concentration of the compounds that will be used to evaluate the sample.

This paper describes a methodology for automating the lane detection in TLC images. After an initial integration of image data onto a one-dimensional profile, lane detection is performed in three phases. In the first phase most of the lanes are detected, some false lanes being eliminated in the second phase. In the last phase a refined search is applied in order to find more difficult lanes which were not identified previously. This paper is organized as follows. Section 2 describes the methodology that was developed for automating the lane detection. The results are presented and discussed in Section 3. Finally, Section 4 is dedicated to the conclusions of this work.

2 Lane Detection

The proposed procedure is applied on the image's region of interest (ROI), previously delineated through an automatic process [3]. The image ROI is initially processed and projected onto the horizontal axis (vertical projection) in order to obtain a one dimensional profile that integrates the data of each lane onto a single dimension. Then, lane detection is performed in three phases. The first phase aims at obtaining an initial set of candidate lanes, which are further validated or removed in the second phase. The third and last phase is a refinement step that allows the inclusion of lanes that are not clearly distinguishable in the profile and that were not included in the initial set.

The ROI detected is converted to grey scale and a closing morphological operator is applied using a square structuring element with a side length of 10% of the image's height. This closing operation allows us to get background information from the image. The grey scale image's ROI is then subtracted from the closed one so that the relevant information, mainly formed by the image bands, is kept.

In the next step a projection onto the horizontal direction of this previous image is obtained. At this point, the top image lines are not taken into account as usually they contain a lot of noise due to the chromatographic process. Moreover, that region is not important for the remaining phases of the method, as all compounds used as markers of FD are associated with the bottom/middle bands. After excluding 25% of the lines starting from the top, the average value of the intensity information from the other lines is calculated to obtain the vertical projection.

Figure 1(a) illustrates an example of a chromatography image and Figure 1(b) shows the ROI (converted to grey scale) obtained with the automatic segmentation described in [3]. In this figure, 12 lanes (vertical tracks) are present. The last 2 are reference lanes while the others represent patient samples. Figure 1(c) is the output of the processes applied to the ROI. The lanes in this figure correspond to the local intensity maxima on the vertical projection that is shown in Figure 1(d).

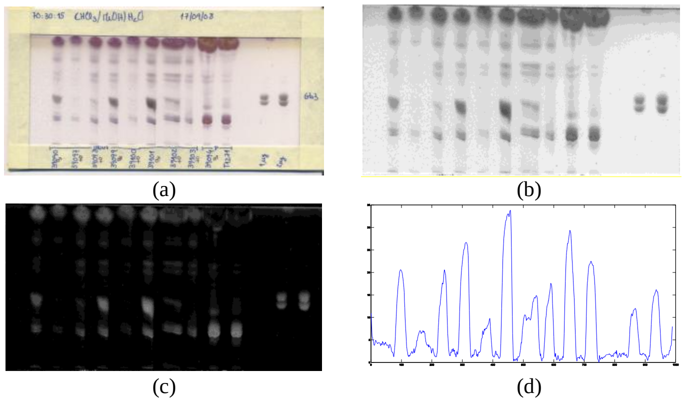


Figure 1: Detection and processing of the ROI to obtain its vertical projection. (a) Original image; (b) Outcome of the ROI segmentation; (c) Result of background elimination; (d) Profile obtained from the vertical projection.

Although the projection operation allows some data integration, the ROI profile still presents some small local variations that make lane detection a hard task. In order to overcome this problem a smoothed version of the profile just containing the main intensity variations is required. The use of common average filters was not able to produce a satisfactory response as they also blurred the more relevant profile transitions. In the proposed approach, a Savitsky-Golay filter was used to smooth the original intensity profile. The Savitsky-Golay filter is based on polynomial regression and it is essentially a weighted average method in the form of

$$g_i = \sum_{n=-n_L}^{n_R} c_n f_{i+n} \tag{1}$$

where n_L an n_R are the number of points used "to the left" and "to the right" of a data point i and g_i is computed as the average of the data points from f_{i-n_L} to f_{i+n_R} , where each point has its own weight c_n, defined by the degree of the polynomial fit [4]. The best results regarding this filter are obtained when the window's width is between 1 and 2 times the FWHM (full width at half of maximum) of the desired features in the data [5].

The lane detection process is performed in three phases with the objectives of: 1) select an initial set of potential lanes; 2) remove false lanes; 3) detect lanes which were not included in the initial set. While the first step is sufficient in most images, the second step is important for noisy and low contrast images, and the third step is essential when the number of bands in a lane is small, as occurs in reference lanes.

In the first phase, the smoothed profile is analysed using two morphological transforms, h-maxima and h-minima. [6], in order to compute the signal's local extremes. Each one of these functions returns 1 if the analysed point is a local extreme of the signal and 0 otherwise. The smoothed signal is searched for regions that were designated as true by the h-maxima function and as false by the h-minima. We get a first set of potential lanes by finding the mean point of each selected region and then label it as the lane's centre.

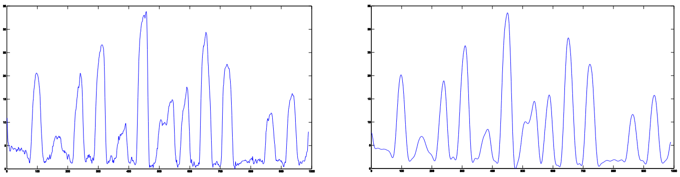
Secondly, we try to remove all the false lanes detected earlier. Each one of the potential lanes is tested against the others for the distance to adjacent lanes, the distance to image borders and the width of the region that originated the potential lane. We expect to detect and remove some false lanes detected earlier due to noise or effects from handwriting on the gel plate.

Finally, the derivative of the signal is analysed. In the profile derivative, a lane is characterized by the occurrence of two local extremes, a local maximum to the left and a local minimum to the right, which can be used for delimitating lane boundaries. After locating the boundaries of all detected lanes, an average lane width can be estimated and afterwards used to try to find new lanes that were not included in the initial set. We search on unoccupied profile regions to check if there is a pattern which coincides with the one found in the lanes. This pattern must also have sufficient amplitude in order to be considered as a lane.

3 Results

The dataset used for testing the proposed algorithm is formed by 66 images. These images with different resolutions and dimensions ranging from 4927×2530 to 569×625 pixels were previously resized to 1024 lines while keeping the lines/columns ratio.

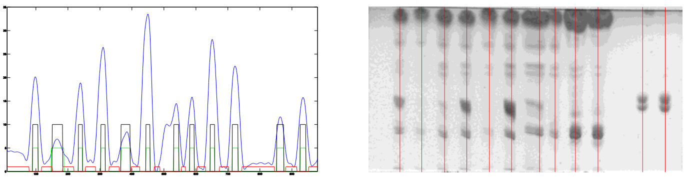
Regarding the application of the Savitsky-Golay filter, as the FWHM of the lanes usually is around 30 – 40 pixels, we decided to fix the filter's parameters as $n_L = n_R = 30$. The filter's order was set to 4 to keep a balance between tracking narrower lanes while still smoothing broader ones. This filter was applied to the profile of figure 2(a) to generate the result depicted in figure 2(b).



(a) (b)

Figure 2: Profile obtained from a ROI of a chromatography image (a) and after being filtered by Savitsky-Golay method (b).

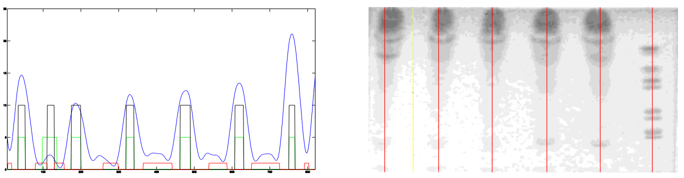
The first phase of the lane detection method allows the establishment of an initial set where the majority of the lanes are present. Figure 3 illustrates the results of this procedure to the original image of Figure 1. Figure 3(a) illustrates the h-maxima and h-minima transforms (logical 1 indicates the presence of a local extreme). In this figure, the black signal indicates the intersection between the h-maxima and h-minima. The red vertical lines in Figure 3(b) represent the detected lanes. The procedure in the first phase was able to correctly identify all the lanes in this image.



(a) (b)

Figure 3: The results of the h-maxima (green), h-minima (red) and the intersection of h-maxima (1) and h-minima (0) in black (a). Set of potential lanes represented by the red lines (b).

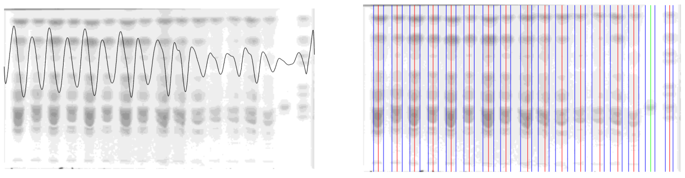
In Figure 4, an example of an image that has a false lane removed by the second phase is shown. The distance between the lane represented by the yellow line and the adjacent lanes lead to state it as a false detection. This distance is considerably smaller than the average distance within the set, as can be verified in Figure 4(b).



(a) (b)

Figure 4: Results of the first phase (a). Red and yellow lines represent the validated and removed lanes after the second phase, respectively (b).

In Figure 5(a) the ROI of an image is represented, with the profile derivative overlapped. The first phase, when applied to this image, missed one lane (the 16th). The analysis of the profile derivative allows us to determine the limits of the lanes and represent them as the blue lines in Figure 5(b). With this technique the lane that was missed in the first phase could now be found - Figure 5(b).



(a) (b)

Figure 5: Profile derivative (a) and result after the third phase (b). The lane detected in this phase is represented in green.

The complete set of 66 ROI's contains a total of 651 lanes. After the 3 phases, our approach allows the automatic detection of 647 lanes. It also finds 20 false lanes. The results obtained are presented in TABLE I.

	True lanes detected	False lanes detected	Lanes missed
Phase 1	644	31	7
Phase 2	644	20	7
Phase 3	647	20	4

TABLE I: Results for the different phases of the method.

4 Conclusions

We proposed a new method for the automatic detection of lanes in chromatography images.

The Savitsky-Golay filter has proved to be a valuable technique to deal with the noise present in all images, smoothing the signal and thus reducing the number of false lanes detected, while still preserving the features of narrower lanes.

The smoothed profile is afterwards used for determining a set of potential lanes based on the detection of local maxima and minima, which are further validated using the distance between lanes and average lane's width. Finally, the profile derivative is analysed in specific areas in order to find missed lanes.

The proposed methodology was successfully evaluated in 66 digital images of chromatographic plates, showing a recall of 99.4% and a precision of 97%.

References

[1] M. Merta, J. Reiterova, J. Ledvinova, H. Poupetova, R. Dobrovolny, R. Rysava, D. Maixnerova, J. Bultas, J. Motan, J. Slivkova, D. Sobotova, J. Smrzova, V. Tesar. A nationwide blood spot screening study for Fabry disease in the Czech Republic haemodialysis patient population. *Nephrol Dial Transplant* , vol. 22, 2007, pp. 179–186.

[2] L. Rodrigues, M. Ferraz, D. Rodrigues, M. Pais-Vieira, D. Lima, R. Brady, M. Sousa, M. Sa-Miranda. Neurophysiological, behavioral and morphological abnormalities in the Fabry knockout mice. *Neurobiology of Disease*, Vol 33(1), 2009, pp.48-56.

[3] A. Sousa, A. Mendonça, M. Sá-Miranda, A. Campilho. Segmentation of the region of interest in chromatographic images. *RecPad2010*.

[4] F. Chau, Y. Liang, J. Gao, X. Shao. *Chemometrics – From Basics to Wavelet Transform*. John Wiley & Sons, 2004.

[5] W. Press, B. Flannery, S. Teukolsky, W. Vetterling. *Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.

[6] P. Soille, *Morphological Image Analysis – Principles and Applications*. 2nd Edition. Springer-Verlag, 2004.

Acknowledgements

This work is being supported by FCT under contract FCOMP-01-0124-FEDER-010913 (Ref^{a}. FCT PTDC /SAU-BEB/100875/2008)