[[1]](#footnote-1)

An automatic detection method of oligoclonal bands on gel electrophoresis of cerebrospinal fluid

S.Boudet, Z. Wang, L.Peyrodie, and G.Forzy\*

*Abstract*— The detection of oligoclonal bands on the cerebrospinal fluid (CSF) electrophorisis is an important diagnostic element for Multiple Sclerosis (MS). The profil images being often highly artifacted and having a very low contrast, the interpretation is often difficult. Here, we propose an automated method of band detection to ease analyze and reduce subjectivity. The method is broken down on multiple steps which each have to be optimized: convert to a grayscale image, realign the profils, remove the artifacts, convert the image to a signal, detect and thresholding the peaks and remove peak which are seen outside the profile. The results of the method on 36 profiles (21 positive and 15 negative) is compared to the blind analysis of an expert biologist. Considering the expert analyze as a ground truth, the method reaches a sensibility of 0.76 and a specificity of 0.80 which is promising view to the difficulty of the problem.

# INTRODUCTION

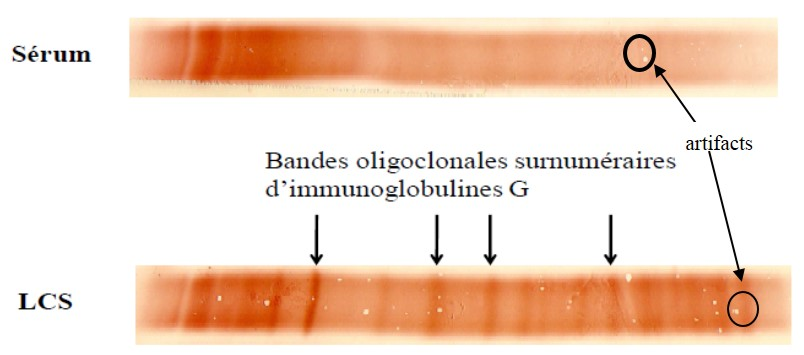
In France, Multiple Sclerosis (MS) affects approximately 1 person in 1000. It is the first non-traumatic cause of severe disability in young patients acquired. It is characterized by inflammatory lesions of the white matter of the central nervous system (CNS), spread in space and time. The paraclinical spatial spread is based on the discovery of 3 of the 4 criteria Barkhof brain and spinal cord on MRI, or on a combination of 2 lesions suggestive MRI associated with a positive cerebrospinal fluid (CSF). Freedman and al [1] recommendations is cerebrospinal fluid (CSF) determinations of immunoglobulin-gamma (IgG) immunoblotting oligoclonal bands utilizing isoelectric focusing (IEF). The electrophoresis is a method that separate molecules on the basis of their size, electric charge and other physical properties **Les IgG sont des anticorps produits dans le LCR des patients. Ces anticorps sont multiples et variables pour chaque patient conduisant à des profils très variables.** **Les bandes de protéines sont révélés par immunoblot avec un chromogène, dont la révélation produit un renforcement là où il y a des protéines en laissant un bruit de fond sur toute la zone de migration des IgG. Le colorant est plus ou moins spécifique des immunoglobulines marquées par une anti-globulinnes. A l’issue de la migration, ce marquage permet de repérer les immunoglobulines G présentes dans le LCS sous formes de bandes. L’image résultante est issue de la superposition du colorant, des bruits additionnels et des protéines contenues dans le LCR. L’intensité de la coloration est proportionnelle à la concentration en protéines de la bande d’IgG. On appelle ligne de base l’interaction du colorant avec le support et cette ligne de base peut être considérée comme un bruit de fond. Nous pourrons alors observer des bandes de fortes ou faibles intensité noyées dans un bruit de fond. Pour interpréter, le biologiste compte les bandes sombres qu’il identifie par leurs aspects fins et homogènes contrastés par rapport au bruit de fond**. Profil is said oligoclonal if there are at least three IgG bands in CSF not present on the serum, (Fig1). **L’enjeu méthodologique est alors triple : pouvoir repérer ces bandes de faible intensité, pouvoir compter le nombre de bandes et pouvoir isoler les artéfacts susceptibles de polluer l’isofocalisation.** It would be a great use to have automated software which realizes the band detection to

This paper aims to propose a multiple step method to automatically detect the oligoclonal bands and to evaluate the concordance with an expert biologist.

# Quality of Gel and Treatment on LCS

**Pour séparer les immunoglobulines G selon leur pHi, nous utilisons la technique Helena d’isofocalisation sur gel d’agarose (Helena Biosciences, United Kingdom).**.

Examples of such gel images are shown in Figure 1, with Fig. 1 this gel image have non-uniforrn backgrounds, noisy stains, and long-tailed smeared bands. [image avec uniquement LCR]



The analyze of profile images is a difficult problem due to the very low contrast and the important variability of baseline level.

Gel electrophoresis images contain various types of artifact or distortion such as blur, or presence of white points (salt and pepper noise), low contrast bands, horizontal lane deformation (smiles). Finally, the resultant signal is a mixture of noise and proteins which reduce the interpretation subjectivity.

Analyze of DNA is the major field of band detection in biology. Several methods [3-5] and commercial interfaces exists. All propose a preprocessing of the gel image captured by camera. Geometry of the whole image is rectified and filtered, then individually each lanes are separated by image segmentation. Because of the quality of image, automatic filtering and rectification could induce some mistake in peaks detection, some existing band could disappears or false detection appear. Detecting gel bands is done individually on the selected lane. To avoid this some semi-automatic techniques have been develop but manual settings are needed such as grayscale threshold selection for a Region of Interest (ROI) or for removing manually specific artifacts. Artifact removing is still a challenge in DNA band detection in addition to being rather time consuming

Use of mixtures models like Gaussian Mixtures [6] or Wavelet transforms [7] is becoming a norm for more accurate peaks detection and for fine tune the quantification of bands on DNA. Automatic parameter setting for mixture model is very difficult and unfortunately, the DNA bands are very more contrasted compared to oligoclonal bands and the methods as set cannot be applied directly on such low quality images. We assume that an optimized preprocessing of image should be useful for both traditional approach enhancement and multimodal ones.

The challenge here is to find all the true oligoclonal bands, forget or add strips could distort the analysis specialist. The aim of the work is to provide an automatic pre-treatment in order to avoid as much as possible misinterpretations.

# Band Detection Methods

## A. Profile image acquisition

The digital image is derived from the physical membrane scanned in 48 bit color at a resolution of 600 dpi [reference du scanner et type de capteur]. The membrane image contains about ten profiles of various patients recorded simultaneously. A specialist has to choose one and select manually a region of interest thanks to a Matlab interface.

## B. Grayscale conversion

## C. Profile re-alignement

Some images are subject to distortion which causes band not being straight and not being horizontal. A re-alignment process is then applied which consist to determine an offset for each column of pixel. This set of offset make a curve and this curve is determined to maximize correlation between each column of pixels while being continue and with minimum distortion.

We define a cost function for this curve:

[]

We will look for the curve which minimizes this cost function thanks to a Monte Carlos method.

## D. Cleaning salt and pepper noise

## E. Obtaining mono-dimensional signal

## F. Filtering and baseline removal

## G. Detect peak and thresholding

## H. Removing peaks which are also detected on white part

# Evaluation

## A Method

## B Result

## C Discussion

sur l’efficacité de chacune des étapes

Scanner, rare mais nous avons déjà observé des cas de bandes qui ne sont plus identifiable après scanner

L’étape de sélection de la zone d’intérêt pourrait également être automatisée mais n’est pas la partie qui représente le plus de difficulté pour le praticien

Cas de distortion non constante

Cleaning salt and pepper noise tres efficace dans l’ensmble mais peut encore être amélioré. il reste un certain nombre de résidu d’artefact qui engendre des fausses detections. Surtout si l’on venait à descendre le seuil.

Obtaining mono-dimensional signal seems conform to the image on most case. Some cases of double peak are merged and sometime the contrast seems less important than on image without knowing exactly why.

Seuillage est sans doute une des pistes devant être le plus amélioré. Nos essais d’utiliser la largeur de bandes ainsi que les valeurs des dérivés n’ont pour l’instant rien donné de significatif donc nous avons préferer rester simple. => Utiliser les GMM pour mieux carracteriser les pics. Combiner de plusieurs parametres avec un apprentissage sur une nombre important de cas.

Removing peaks which are also detected on white fonctionne assez bien, quelques faux positives et faux negatives tout de meme. Toutefois dans certains cas les experts considèrent qu’une bande est un artefact du fait qu’elle se répète aux profils voisins (pas seulement sur la zone blanche). Réaliser ce proceder automatiquement peut être difficile.

# Conclusion

Confrontation between ground thruth and automated reading give promising results. We have shown how by improving and adapting the treatment at the profile it is possible to improve the peak detection with conventional methods. The ability to automatic filtering artifacts contributes to improving the efficiency of results. Our future work therefore focuses on the reliability of the detection of peaks by methods based on mixture models.

Appendix

Appendixes should appear before the acknowledgment.

Acknowledgment

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression, “One of us (R. B. G.) thanks . . .” Instead, try “R. B. G. thanks”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

References

1. Freedman and al. “Recommended standard of cerebrospinal fluid analysis in the diagnosis of multiple sclerosis: a consensus statement,” in *Arch Neurol*, vol 6, june*,*2005, pp865-70.
2. Delpech B, Lichtblau E. “Immunochemical estimation of IgG and albumin in cerebrospinal fluid”,*Int J Clin Chem*; vol 37: mars 1972, pp 15‑23.
3. N. Kaabouch and R Schultz,"A 2-D gel electrophoresis DNA image analysis algorithm with automatic thresholding," *The international Society for Optical Engineering Proceedings (SPIE)*, Vol. 6508,2007, pp. 65081H-I to 65081H-12,
4. A PSU, A Bhornra, "Automatic analysis of agarose gel images," *Bioinformatics* ,vol 17(11), 2001,pp. 1084-1090,.
5. I Bajla, L Hollander, S. Fluch, K Burg, Kollar, "An alternative method for electrophoretic gel image analysis in the Ge!Master software," *Computer Methods and Programs in Biomedicine* , vol 77, 2005, pp. 209 -231,.
6. Polanski A, Marczyk M, Pietrowska M, Widlak P, Polanska J (2015) Signal Partitioning Algorithm for Highly Efficient Gaussian Mixture Modeling in Mass Spectrometry. PLoS ONE 10(7): e0134256. doi:10.1371/journal.pone.0134256.
7. B. Moreira, A Sousa, A Mendonça, and A Campilho, “Automatic Lane Segmentation in TLC Images Using the Continuous Wavelet Transform,” Computational and Mathematical Methods in Medicine, vol. 2013, 19 pages,. doi:10.1155/2013/218415.

1. \*Research supported by Conseil Régional NPDC.

   S.Boudet. is with FMM-UTSB, 56 rue du Port, 59000 Lille (France) (corresponding author to provide phone: 00330328384858 fax: 00330328384838; e-mail: samuel.boudet@gmail.fr).

   Z. Wang., is with FMM-UTSB, (e-mail: zefeng.wang@hotmail.fr).

   L.Peyrodie is with UTSB-HEI, 13 rue de Toul, 59000, Lille (e-mail: [Laurent.peyrodie@hei.fr](mailto:Laurent.peyrodie@hei.fr)).

   G. Forzy is with GHICL, rue du Grand But, BP 249, 59642 Lomme, (France). [↑](#footnote-ref-1)