Automated image analysis of gel electrophoresis of cerebrospinal fluid for oligoclonal band detection

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

***Abstract*— Detection of oligoclonal electrophoretic bands in cerebrospinal fluid (CSF) is an important diagnostic tool for Multiple Sclerosis (MS). Electrophoretic profiles are difficult to interpret due to low contrast and artefacts. A semi-automated method to ease analysis and to reduce subjectivity is presented. The method sequentially converts color images to grayscale, realigns bands, removes artifacts, then converts 2D images to a signal, before detecting, thresholding and editing peaks to optimize profiles. Such treated profiles (21 positive and 15 negative) are compared to ground truth analysis of an expert biologist. 16 profiles over 21 are well detected positive and 12 profiles over 15 are detected negative, results seem similar to inter-experts variability reported in literature.**

I. Introduction

In France, Multiple Sclerosis (MS) incidence is one per 1000 and is the leading acquired non-traumatic cause of severe disability in young patients. Inflammatory lesions of white matter in the central nervous system (CNS) spread in space over time. Diagnosis and disease severity are based on MRI evidence for 3 out of 4 Barkhof brain and spinal cord criteria or a combination of two suggestive MRI observed lesions associated with a positive cerebrospinal fluid analysis (CSF) [1]. Immunoglobulin-gamma (IgG) immunoblotting of cerebrospinal fluid (CSF) by isoelectric focusing (IEF) is a diagnostic tool[2].

Multiple and varied IgG antibody electrophoretic profiles in CSF from patients are observed. Chromogenic immunoblotting accentuates protein coloration over background and improves marking, superposing coloration. Color intensity and IgG concentration are proportional. A base line or background noise is determined as chromogen interaction with native immunoblot supports. Bands of varied intensity are subject to this backdrop. The clinician distinguishes bands, as thin and homogeneous prior to counting. The presence of at least three IgG bands refers to an Oligoclonal Bands (OCB)-positive sample. As reported in [3], there is several possibility of misinterpretation due to blot and stain determination difficulties. This create a lack of reproducibility which could lead to false-negative/positive in critical CSF samples, i.e., samples with few and weak bands. Diagnostics thus depend on the detection of weak bands, the elimination of artefacts and accurate counting.

*Research supported by PHRC POLAR.

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).

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

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

G. Forzy is with FMM and GHICL; rue du Grand But, BP 249, 59642 Lomme, (France).

A semi-automated process would thus be less subjective, simpler in analyses and computation. Several methods [4-7] and interfaces exist for DNA gel analysis, using camera captured images. Geometry is rectified and filtered and each individual lane segmented. These methods including Gaussian Mixture models [8] and Wavelet transformation [9], used to fine tune quantification of DNA bands can be disrupted by artifact. Semi-automatic techniques for Region of Interest (ROI) or removing specific artifacts also exist. Nonetheless, analyses are still dependent on manual grayscale threshold selection.

In contrast for IgG isofocalisation, to count oligoclonal bands does not require fine band detection. However contrast is much lower and band can be difficult to perceive by the naked eye. Any small perturbation can thus cause false detection. For those reasons, the method to process DNA electrophoresis cannot be applied directly on IgG isofocalisation without adapting methods and parameters. We did not find any reference on literature [10] which deals about this problem. Such a process would need to overcome several difficulties:

- Low to signal noise ratios: bands are often of low amplitude.
- Image artifacts: black and white stains referred to as salt and pepper noise.
- Blurs: false lanes over the entire blot membrane.
- Smiles: horizontal lane deformations.
- Baseline variation.
- Inter-blot differences of baseline.

This paper reports on a multiple step method to read IgG oligoclonal bands and to evaluate concordance with an expert diagnostician.

II. Band Detection process

The detection process steps on a high contrasted positive profile with some artifacts and smile distortion are illustrated (Fig 1, A to E and Fig.2).

A. Electrophoresis

IgG electrophoresis methodology was according to the supplier of agarose gel and immunoblot membrane (10 cm x 8cm, Helena Biosciences). The image contains ten profiles of various patients recorded simultaneously.

B. Profile image acquisition

The physical membrane is scanned [Epson V750 PRO] in 48 bit color (3 color values are scaled from 0 to 1) at a resolution of 600 dpi. A specialist selects a Rectangular Region of Interest (ROI) assisted by a specific interface (Matlab). (Fig. 1.A-B).

Preprint submitted to 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Received March 10, 2016.

Figure 1. Image processing steps – A. entire lane on the membrane with its ROI – B. Annotated profile image of ROI – C. Grayscale conversion and re-alignment – D. Noise removal. E. Bands representation on original image

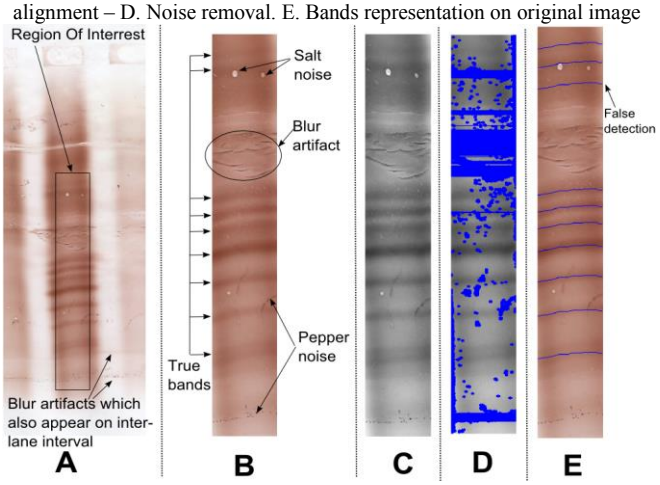
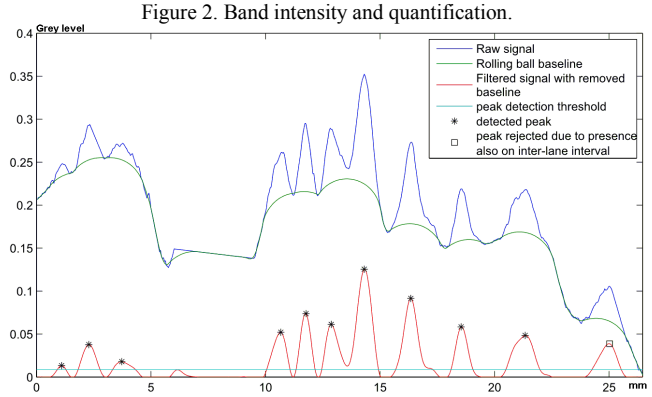


Figure 2. Band intensity and quantification.



C. Grayscale conversion

Conversion to grayscale is linear (Gray=R*Red +G*Green +B*Blue). In order to find the coefficients which maximize the contrast, a Principal Component Analysis (PCA) has been realized on some positive profiles (considering all pixels as samples and the three colors as data). The coefficients obtained from various profiles are normalized and then averaged. The final coefficients obtained for all images are R=0.16, G=0.52, B=0.32. (Fig. 1.C).

D. Band realignment

Some images are subject to distortion called smiles and bands may not being perfectly straight and 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 t_i $(i = 1 \dots n)$ and this curve is determined to maximize correlation between each column of pixels while being smooth and with minimum distortion.

Let I being the image, I_k being the signal of the k-th column of pixel. Eq.1 designs a cost function $C(t, I)$ which has to be minimized by an optimal t.

$$C(t,I) = \alpha C_{\text{distortion}}(t) + \beta C_{smooth}(t) + C_{corr}(t,I)$$

$$C_{\text{distortion}}(t) = \|t\|_1 \quad (1)$$

$$C_{smooth}(t) = \|d^2 t\|_1$$

$$C_{corr}(t,I) = -\sum_{i,j\ (i\neq j)} r^2(I_i * \delta_{t_i}, I_j * \delta_{t_j})$$

$\|.\|_1$ correspond to the norm 1 of the signal (sum of absolute value), d^2t is the second derivative of t, $r^2(.,.)$ is the squared correlation between the two signals, $I_i * \delta_{t_i}$ is the i-th column of the image shifted by t_i pixels, α and β are balancing terms set empirically.

This cost is three balanced terms:

- $C_{\text{distortion}}$ (sum of absolute values of the curve) aims to minimize the change when no significant distortion is present;
- C_{smooth} aims to minimizes the second derivative d^2t of the curve to avoid discontinuities while allowing not horizontal but straight shifting.
- C_{corr} Negative of the sum of the correlation of each pair of shifted pixels.

The optimization process uses a greedy algorithm: each column I_i of pixels is individually treated in random order, optimal shifting t_i is determined by testing all values over the range taking into account the current shifting of other columns. This process is repeated several times so that each column is optimized 4 times. This process is programmed in C and it takes a neglecting computation time (<0.5s by profile). A result is illustrated on Fig. 1-C.

E. Removing salt and pepper noise

This part aims to reject the artifacted parts on the image in order to get a clean signal when averaging horizontally.

For each pixel p of grey value I_p, we compute the average A_p of an ellipse of 10 pixel wide and 5 pixel height centered on p. If the absolute difference between I_p and A_p is over a threshold, the pixel is considered as artifact as well as the other pixels around on a radius of 3 pixels. Ellipses of larger width than height avoid rejecting contrasted horizontal bands. This rejection process is set to be highly sensible and poorly specific since the information of a band can be retrieve on all the width of the profile.

At the end, rows of pixels of more than 50% artifacts pixels are completely rejected (Fig. 1-D).

F. Conversion of 2D image to mono-dimensional signal

Upon realignment and cleaning, a single dimensional signal is obtained by averaging each row of the profile image. Missing values occur due to rows rejected during the previous step. If less than 15 pixels (0.65mm on membrane) occur between two missing values, the signal between them is also rejected. Blocks of missing values are filled in through linear interpolation at the borders (Fig. 2 in blue).

The aim now is to detect peaks on the signal which are the local maxima.

G. Filtering and baseline removal

A low pass filter is applied on the signal to remove small noise and to have smooth curve with reliable local maxima. The filter is a 4-th order Butterworth with a cutoff period of 0.67mm (16 pixels) applied with a forward-backward process to avoid phase-shifting.

Then, signal baseline is removed using a rolling ball algorithm [7] (Fig. 2 in red) with radius of 1.48mm on distance and 0.041 on grey level (=10.5 at 256 full scale).

This rolling ball has three interests: amplifies peak which are on slopped baseline, remove the peaks which are two wide to correspond to a lane and ease the amplitude determination of a peak.

H. Peak detection and threshold setting

Upon filtering and baseline removal, the peak are defined as the local maxima above a threshold, set to 0.0087 (=2.21 at 256 level). This threshold is near the limits of human eye perception (Fig. 2 in cyan).

Close local maxima can appear above the threshold without returning to the baseline. It can either correspond to two bands which are very close or just a shape effect. When the minimum between such peaks is over 2/3 of the lower of the two peaks, this peak is removed.

I. Removal of inter-lane distortion

Blurs can have the same shape as bands and produce false positives. Most of time, these artifacts appear on the inter-lane interval on each side of the profile. The detection process (steps F, G H) is then applied on the white image of inter-lane interval. The images taken correspond to an 11-pixels wide rectangle in the middle of the inter-lane interval each side of the profile. When bands appear either to the left or to the right, all bands in the profile within a 15-pixel range are considered as blur artifacts and removed (Fig. 2 squares).

III. EVALUATION

A Method

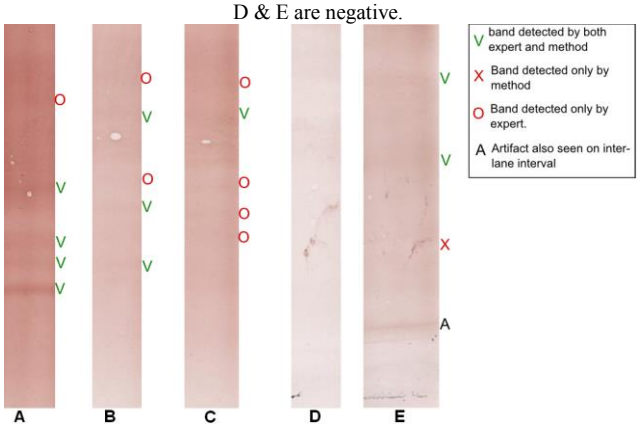
To evaluate the method, a blind comparison with the analysis of an expert biologist has been realized. Our database is composed of 36 profiles (21 positive and 15 negative) on 5 different membranes. In a first step, each profile is analyzed and classified as oligoclonal or not by a consensus of 3 experts, directly on membrane.

In a second step, an expert biologist has selected of ROI on scanned profiles and has reported precisely the bands by mouse thanks to a specific Matlab interface. The profile is then analyzed by the described method and the two analyses are compared. Bands positioned by expert and computer were considered to match when less than 15 pixels (0.65mm on membrane) separated them, retaining the nearest bend when multiple matches occurred.

B Result

The results of five profiles are illustrated (Fig. 3). A, B and C are positive profiles and show 5 bands each. They have decreasing contrast which leads to decreasing performance. On profile A, the contrast is relatively important, and only one low amplitude band is missed by the method. Profiles B and C are positive with low contrast. They illustrate the difficulty of the problem for both eye and computer analysis. B is computer detected positive with three bands whereas C is detected negative with only one band. Profiles D and E are negative with respectively 0 and 2 bands. On profile D, the artifact is well ignored and no band is detected. On profile E, the method well detect the 2 bands but add a false detection which make this profile being a false positive.

Figure 3. Result of band detection on 5 profiles: A, B & C. are positive; D & E are negative.



Considering the expert analyze as a ground truth, we mean by "well classified (vs misclassified)" an agreement (vs disagreement) between expert and automatic analysis. Results are summarized as follows (Table 1.).

TABLE I. RESULTS OF PROFILE ANALYSIS

Profile	*Well classified*	*Misclassified*	*Matching bands*	*Forgotten bands*	*Added bands*
Positive (21)	16	5	103	37	19
Negative (15)	12	3	7	9	17
Total (36)	28	8	110	46	36

For entire profile classification, we measure a sensitivity of 0.76 and a specificity of 0.80. For band detection alone, it has a sensitivity of 0.67 and a Positive Predictive Value (PPV) of 0.75. Pixel specificity defined as follow (eq. 2) is 0.91 and Pixel Negative Predictive Value (eq. 3) is 0.89.

$$\text{Pixel Sp.} = \frac{\text{Concordant negative pixels}}{\text{Concordant negative pixels+30 pixels per added bands}} \quad (2)$$

$$\text{Pixel NPV.} = \frac{\text{Concordant negative pixels}}{\text{Concordant negative pixels+30 pixels per forgotten bands}} \quad (3)$$

C Discussion

Observed results are similar to the inter-expert variability reported in [3]. The method reaches a good specificity which shows that most of possible false detections have been well avoided. The sensibility is a bit inferior due to very low contrast on some image. By observing the a posteriori results of each step, it seems that each step is efficient enough but there is still a small loss of information which cumulated reduce the classification rate:

- Step B: although scanning enables the user to zoom and maximize contrast, bands are often more directly visible on the membrane. The ROI definition could be automated, but this is not a priority for diagnosticians.

- Step D: band re-alignment worked well, but may be limited to cases where distortion is regular over the profile.

- Step E: Noise removal is efficient. It does never remove true peak on our study, but border residues are a source of false positives.

Preprint submitted to 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Received March 10, 2016.

- Step F: Mono-dimensional signals seems conform to images but there is few cases where double-peaks merge and few cases where the contrast on signal is reduced.

- Step G&H: Low-pass filtering, baseline removal and thresholding should be the principle steps to be improved. Most of forgotten peak can be observed on curve but are under the threshold.

- Step I: Peak removal from inter-lanes interval analysis is efficient. This process should be extended for bands which appear on every profiles of the membrane.

IV. Conclusion

The present results, point the way to the potential of semi-automatic detection analysis of IgG oligoclonal bands on immunoblots. We hope that this work could be used as basis for future research to optimize independently each step. Particularly, our future work will focus on the reliability of the detection of peaks by methods based on mixture models. We hope also testing the method on a more important profile variety and to compare the results with the inter-expert and intra-expert variability in order to have a better idea of true discordances and the questionable ones. The final aim is to propose a tool to ease and improve reproducibility of OCB analysis.

Acknowledgment

We thanks Christophe Herlin for is technical assistance Doctor David Perry for English review and constructive discussion.

References

[1] Polman and al. “Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria”, *Ann Neurol*, 69(2), Feb 2011; pp 292–302.

[2] 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.

[3] Franciotta D, Lolli F.”Interlaboratory Reproducibility of Isoelectric Focusing in Oligoclonal Band Detection”, Clin Chem 53(8), 2007;pp 1557-8.

[4] 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,

[5] A PSU, A Bhornra, "Automatic analysis of agarose gel images," *Bioinformatics*, vol 17(11), 2001, pp. 1084-1090.

[6] 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.

[7] I, Lazar, "Gel Analyzer 2010a: Freeware 1D gel electrophoresis image analysis software".

[8] 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.

[9] 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.

[10] A. Gillain, A. Fumal, C.Nève, J.-M. Minon , A. Maertens de Noordhout,’Oligoclonal bands and IgG index interpreted according to Reiber in the inflammatory diseases of the central nervous system’ Immunoanalyse et Biologie Spécialisée, vol 21(6), 2006,pp 348-356