TECHNICAL BRIEF

A method based on bead flows for spot detection on 2-D gel images

Olivier Langella and Michel Zivy

CNRS, UMR 0320/UMR 8120 Génétique Végétale, Gif-sur-Yvette, France

A method has been developed to detect spots on 2-D gel images. It is based on an analogy with beads flowing uphill on the surface of the gel image and on the analysis of their paths. This method does not need image smoothing, thereby allowing for conservation of the resolution of the original image. In addition, it allows for the detection of overlapping spots, even when there is no valley between their centres.

2-D PAGE allows the study of individual components of complex mixtures of proteins, that are revealed as spots on 2-D gels. Spot detection is a very important issue in the process of automatic analysis of 2-D patterns; not only the accuracy of presence/absence variations between gels, but also the quantification of detected spots depends on this step. Ideally, a simple detection of local maxima should be sufficient. But the problem is more complex: images contain high frequency noise and small artefacts, giving rise to numerous local maxima that do not correspond to spots and to several local maxima within a given spot. High frequency noise can be caused by the image acquisition device or by the grain of autoradiographs. Microheterogeneities can also be caused by CBB particles, silver stain, dust, small bubbles underneath the gel during the scan and limits of broken gels. Thus, a good method of spot detection must (i) detect real spots and (ii) not detect artefactual peaks.

Published methods for spot detection are based on successive thresholding of the image, computation of image derivates, Laplacians, watershed algorithm or direct detection of local maxima separated from each other by a minimum distance [1–11]. These methods are sensitive to high frequency noise, and in general the image is smoothed before spot detection. But smoothing decreases image resolution, to the detriment of the detection accuracy for small or overlapping spots. The method presented here has been developed in view to avoid smoothing before spot detection, and to allow the detection of overlapping spots even when there is no valley between their centres.

The point of the method is that detection is based on a surface criterion, and not on an intensity criterion. Surface computation is based on an analogy with beads flowing on the surface of the gel image. It is useful to consider the 2-D gel image as a landscape, in which high intensities represent high altitudes. In this landscape, spots look like hills (peaks with large bases), and artefactual peaks look more or less like posts (peaks with small bases). Let us consider that one bead is placed on each pixel of the image, and that beads are attracted by high altitudes: they will roll uphill. Every bead will move pixel by pixel, always going in the direction of maximum slope, up to an equilibrium point. All the beads whose starting point was on the slope around a given peak will stop at the top of this peak. Thus, if the peak has a large base (hill), numerous beads will arrive at its top. On the other hand, a small number of beads will arrive on top of a peak with a small base (post), whatever its height. In other words, the number of beads present at the top of a peak is equal to the surface of its base. Thus, one can count the number of beads at each position and detect spots as a function of their surface. In addition, one can keep track of the paths followed by beads. If two spots overlap in such a way that only one local maximum is left, all beads of the region will finish their

Correspondence: Dr. Michel Zivy, UMR de Génétique Végétale, plateforme PAPPSO, INRA − Univ Paris-Sud − CNRS − AgroParisTech Ferme du Moulon, F-91190 Gif-sur-Yvette, France
E-mail: zivy@moulon.inra.fr
Fax: +33-1-69-33-23-40

course at the top of the largest spot. But, if the smallest spot is still distinct from the largest one, paths followed by beads originating from the smallest spot will merge with each other before flowing in the direction of the largest spot. Thus, these overlapping spots can be detected by detecting path confluences.

As images contain noise, there are generally several local maxima on top of a given spot. Consequently all beads of a spot will not reach the same pixel, and the number of beads that arrive at each local maximum is smaller than the total surface of the spot. The same is true for path confluence in overlapped spots: image noise leads to a series of two paths confluences rather than a single confluence of many paths. Thus, spot detection must be based on the selection of groups of close pixels rather than on single pixels. To do this, pixels selected according to their number of beads or confluence paths are used to create an image of spot position probabilities (see below), that will produce a single maximum in regions where they are close to each other. Spot centres are detected on this image, and quantification is performed by allowing beads to transport the intensity of their departure pixel to spot centres.

The program is composed of successive procedures. Procedure outputs are images (same size as original image), that are used as input by the following procedures. In the following, italics are used for parameters that can be modified by the user:

(i) Creation of the DIRECTIONS image: the direction of maximum slope is computed at every pixel of the original image (Fig. 2A). Slopes are computed in the eight directions as follows:

Slope (current) = [*I*(next) − *I*(previous)]/distance

$I(x)$ is the intensity at position x. Distance is equal to 1 in x and y directions and to $\sqrt{2}$ in diagonal directions. Not taking the intensity of the current pixel into account allows beads not to be stopped by small one pixel-large peaks, without loosing faint variations that would be lost by median filtering (Fig. 1A). The produced DIRECTIONS image is an image of codes, each of them indicating the direction to follow at every pixel position (Fig. 2B).

(ii) Creation of the BEADS image: assuming a starting state with one bead *per* pixel, all beads roll to the neighbour pixel indicated by the DIRECTIONS image at their current position, until they reach their stable positions (local maximum). One (one bead) is added at the final position of every bead. Thus, pixels of the BEADS image contain the number of beads arrived at their positions. Beads concentrate in local maxima and disappear from slopes (Fig. 2C).

(iii) Creation of the PATHS image: beads roll again from their initial to final position, but, this time they add 1 to all pixels on which they roll. Thus, every pixel of the PATHS image contains the number of beads that rolled on it (Fig. 2D).

(iv) Creation of the SELECT image: this image will contain possible positions of spot centres. To be selected, a pixel must meet one of the two following conditions: (a) it is a local

maximum in PATHS and at least *minbead* beads arrived at this position; or (b) it is a position of path confluence, *i.e.* at least *minpath* paths used by at least *minflux* beads merged at this position (Fig. 2E). Only selected pixels have a non-null value in the SELECT image. The value is equal to the number of beads arrived to the pixel, or to the *merging bonus* given to the detection of path confluence.

(v) Creation of the PROBABILITIES image: a double Gaussian curve (a bell) is added at every position containing a non-null value in SELECT. The height of the bell is equal to the value found in SELECT. When two pixels selected in SELECT are close to each other, the addition of their double Gaussian curves in the PROBABILITIES image will produce a single local maximum or two distinct maxima, depending on the distance and on bell SDs *sx* and *sy* (Fig. 2F).

(vi) Spot detection: a list of spots is created by selecting the position of local maxima above *minproba* in the PROBABILITIES image. *Minproba* can be greater than *minbeads*: this allows the selection of local maxima at step (iv), which will contribute to spot detection at step (vi) only if they are close to other local maxima (multiple maxima within a spot, Fig. 1B). Thus, spots containing several local maxima can be detected as a single spot without over-detection of small artefacts. At this step, spots are characterized by their x and y positions and by a spot number.

(vii) Spot quantification: quantification is then carried out by determining pixels that belong to the same spot and by integrating their intensities. A pixel belongs to a spot if the bead originating from this pixel finishes its course in the central region of this spot. A NUMBERS image is created in which spot central regions are first formed by labelling the neighbourhood of each spot with its own number. The user can control the size of these central regions and the possibility of merging spots that would still be detected too close to each other. Beads roll once again from every position of the image, but this time they stop as soon as they enter a region labelled by a spot number in the NUMBERS image. Their pixel of origin is labelled with this number. To include spot regions that would not be labelled during this process because of staining artefacts, labelled regions are then enlarged by aggregating neighbour pixels whose value in the original image is lower than that of the border of the labelled region. This process is repeated *npass* times. It makes negligible changes for most spots, except for saturated or diffuse spots that contain several peaks that are too small to be detected and too far from each other to be merged during the building of the PROBABILITIES image. Possible holes within labelled spot regions are also filled in. At the end of this process, the NUMBERS image contains spot surfaces labelled with their numbers (Fig. 2G). Spot quantification is then carried out by attributing original pixel intensities to spots as a function of their label in the NUMBERS image. Local background is estimated according to spot minimum values. An image showing spot boundaries is produced (Fig. 2H).

Figure 2I shows that two overlapping spots that did not produce a local maximum were correctly detected. To test
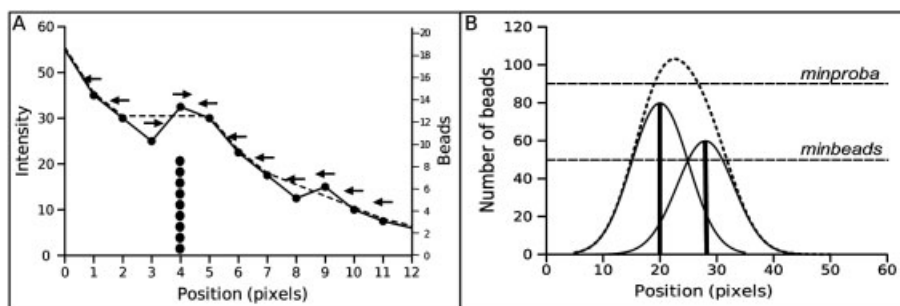
Figure 1. 1-D illustration. (A) Directions computed at each position are symbolized by arrows. They are computed as a function of the intensity (solid line) of neighbour pixels. Nine out of the 11 beads that are visible on this scheme will end up at position 4, because opposite directions are indicated on pixels 4 and 5. This escarpment will thus be detected by the beads method. A median filter (dashed line) would eliminate this local maximum. (B) Gaussian curves were synthesized at positions 20 and 28, because the number of beads that accumulated at their respective positions were greater than *minbead*. Their addition produced a single maximum (dashed line), which will lead to the detection of a single spot since it is greater than *minproba*. Isolated Gaussian curves of same heights would not lead to spot detection.
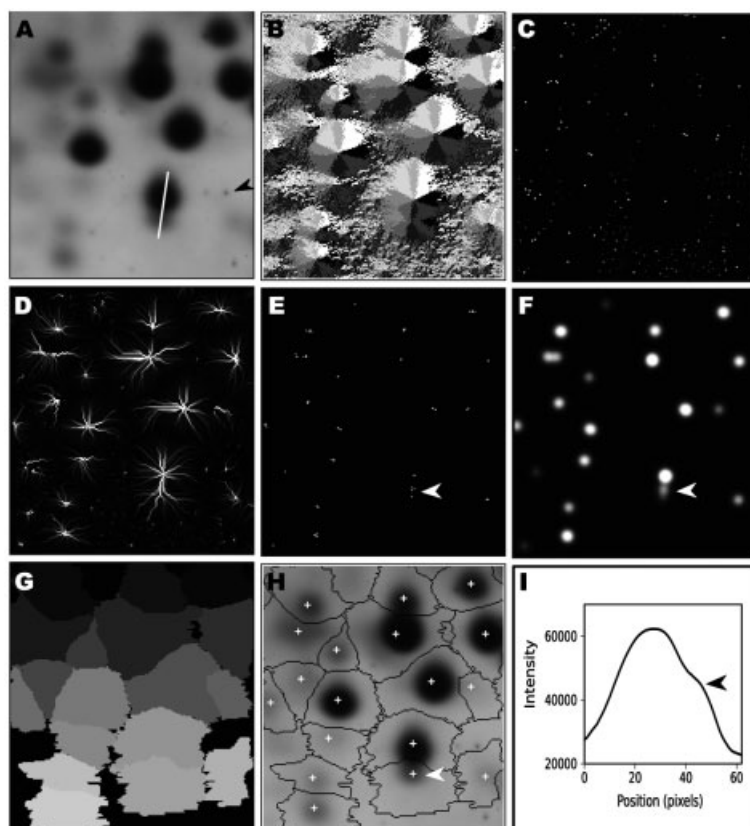

Figure 2. Steps of spot detection are presented on a small part of a silver-stained gel. (A) Original image. The arrowhead shows an artefact that will not be detected although it is more intense than some detected spots (see H). (B) DIRECTIONS image. (C) BEADS image. Beads accumulated at the top of hills. (D) PATHS image. (E) SELECT image. (F) PROBABILITIES image. (G) NUMBERS image. (H) Spot boundaries. The arrowheads in E, F and H show a position where a spot is detected resulting from the detection of path confluences. Beads did not accumulate at this position (see C). (I) Intensity profile along the white line drawn in A: the spot shown by arrowheads in E, F, H and in this profile does not correspond to a local maximum in the original image.

more extensively the ability to detect overlapping spots, we used synthetic images, where 2-D Gaussian curves were used as models for spots. Figure 3A shows that the BEADS method actually allows the detection of spots showing a large overlapping, and that the smallest distance that still allows the detection of both spots depends on their size and intensity. It also shows that the method can tolerate high values of high frequency noise in the image, since the detection of

overlapping spots was only slightly affected by large modifications of pixel values. The ability of the method to exclude artefacts is illustrated in Fig. 3B: no spot was detected in the image of an empty part of a real gel, on which numbers were manually written. This was expected, since the BEADS method excludes contrasted objects by construction.

As it is not possible to explore all combinations of parameters (gel resolution, spot shape, background variations)
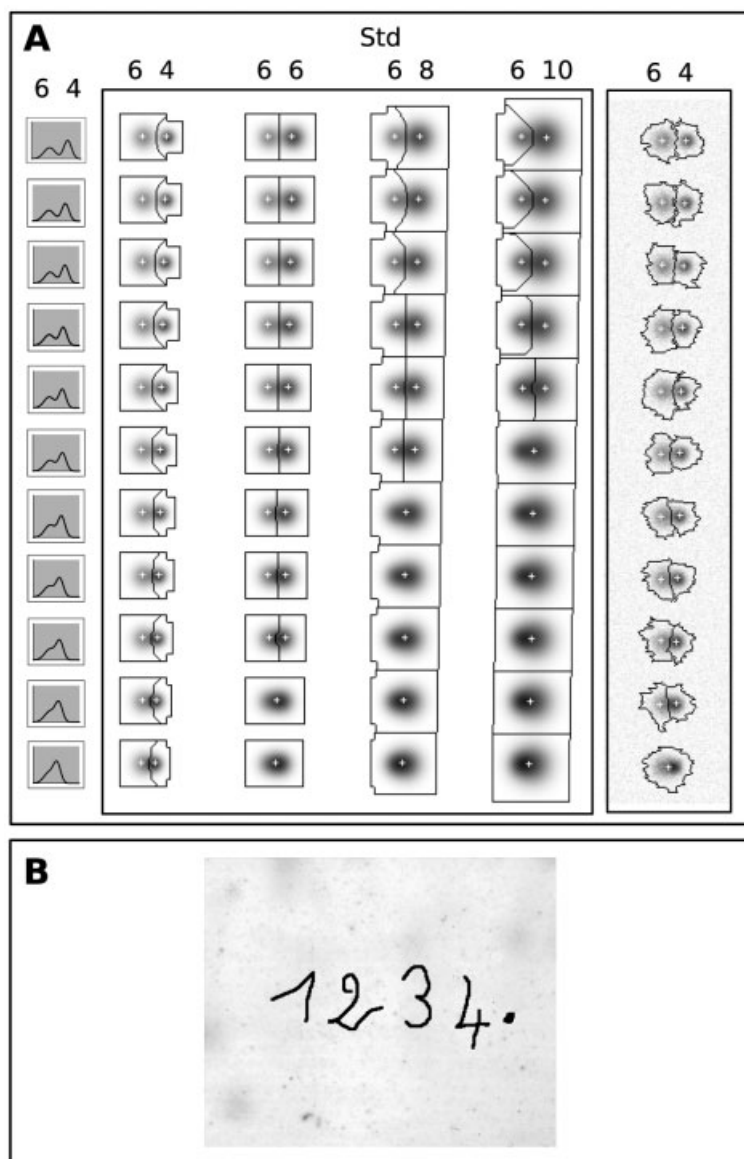
Figure 3. Effects of spot overlapping and background noise on spot detection. (A) A synthetic image was obtained by using 2-D Gaussian models. The distances between members of the same couple of spots decreased from the top (distance = 19) to the bottom (distance = 9) of the image. The intensity of both spots was constant all over the image, (respectively 20 000 and 33 000 for the left and right spots), and the SDs of the right spot varied from 4 to 10, while the SD of the left spot was constant. The left column of (A) shows intensity profiles of the left column of the synthetic image. Boundaries and centres of detected spots are shown. Two spots were detected in all couples of the left column, while the three couples at the bottom showed only one local maximum. The right column of (A) shows the effect of high frequency noise on spot detection. An image containing the same synthetic spots as the left column of the first image was submitted to noise, *i.e.* pixel values were allowed to change around their original value according to a Gaussian model with a SD of 2000. This treatment affected only the detection of the couple where the distance between spots was the smallest. (B) Detection was tested on a part of real gel image containing no visible spot, on which numbers were written. No spot was detected.

that can affect detection accuracy, we tested the method on ten synthetic images of gels parts in which 2-D Gaussian curves were randomly generated (S1 of Supporting Information). In the absence of background, the BEADS method was very similar to the human eye (68 spots detected in average). The addition of a real background lead to the loss of six spots in average, and to a small number of artefactual detections (0–2 *per* gel). Progenesis discovery (PD, Non-Linear) was tested on the same images. In the absence of background, PD missed more spots than BEADS. However, the addition of background had less effect on PD than on BEADS, and both programs performed about the same way regarding the number of 'actual' spots detected. Interestingly, both programs detected spots that were missed by the other (respectively 2.6 and 1.9 spots *per* gel detected only by

BEADS and PD), which shows that the methods employed could be complementary. On the other hand, PD produced more false detection than BEADS. Consistently, many spots were detected by PD in the image shown in Fig. 3B (S2 of Supporting Information) on which BEADS detected no spots. The comparison between the detections of PD and BEADS on a real gel silver-stained gel (S3 and S4 of Supporting Information) confirmed this large difference regarding false detections.

Detection failures of BEADS (spot not detected or split in several pieces) are principally observed for diffuse or saturated spots, *i.e.* in situations where beads belonging to a same spot do not concentrate enough in the central region before stopping: either beads are not numerous enough at their arrival position to be detected, or they are detected too

far from each other to produce a single maximum in the PROBABILITIES image. These situations are more frequent in silver-stained than in CBB-stained gels (S3 and S5 of Supporting Information).

All intermediate images can be saved. This allows the user to examine them by using, *e.g.* ImageJ (http://rsbweb.nih.gov/ij/), to adapt parameters according to observed values. Once parameters have been set, the program can be run without saving any intermediate file. The output can be in the form of images showing spot contours and centre (TIFF, JPEG or SVG format) or in the form of text containing the parameters of detected spots in two possible formats, *i.e.* tabulated text file or PROTICdbML xml format. The PROTICdbML format allows direct uploading to a PROTICdb database [12], where the image can be displayed, and various information can be attached to detected spots. As the BEADS program does not include interactive and matching tools, it cannot be used by itself for comparative analysis of several gels.

The BEADS program has been developed in C++, with the aid of the Cimg library (The C++ Template Image Processing Library, http://cimg.sourceforge.net). Any format of image readable by ImageMagick can be used, although lossy compression will result in deteriorated results. The program has been tested mainly on images of gels scanned with a 300 ppp resolution and 16-bit TIFF images. The entire process takes about 20 s of which 10 s are used for loading the original image on a Pentium D 3.2 GHz processor for the detection of 2200 spots on a 3000×2500 pixels image. The program is open source and freely available at http://moulon.inra.fr/beads/beads.html.

The authors have declared no conflict of interest.

References

[1] Anderson, N. L., Taylor, J., Scandora, A. E., Coulter, P., Anderson, N. G., The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns. *Clin. Chem.* 1981, *27*, 1807–1820.

[2] Appel, R. D., Palagi, P. M., Walther, D., Vargas, J. R. *et al.*, Melanie II-a third-generation software package for analysis of two-dimensional electrophoresis images: I. Features and user interface. *Electrophoresis* 1997, *18*, 2724–2734.

[3] Bettens, E., Scheunders, P., Van Dyck, D., Moens, L., Van Osta, P. *et al.*, Computer analysis of two-dimensional electrophoresis gels: A new segmentation and modeling algorithm. *Electrophoresis* 1997, *18*, 792–798.

[4] Bossinger, J., Miller, M. J., Kiem-Phong, V., Geiduschek, P., Xuong, N. H., Quantitative analysis of two-dimensional electrophoregrams. *J. Biol. Chem.* 1979, *254*, 7986–7998.

[5] Cutler, P., Heald, G., White, I. R., Ruan, J., A novel approach to spot detection for two-dimensional gel electrophoresis images using pixel value collection. *Proteomics* 2003, 3392–3401.

[6] Garrels, J. I., Two-dimensional gel electrophoresis and computer analysis of proteins synthesized by clonal cell lines. *J. Biol. Chem.* 1979, *254*, 7961–7977.

[7] Lipkin, L. E., Lemkin, P. F., Database techniques for multiple two-dimensional polyacrylamide gel electrophoresis analyses. *Clin. Chem.* 1980, *26*, 1403–1412.

[8] Pleissner, K. P., Hoffmann, F., Kriegel, K., Wenk, C. *et al.*, New algorithmic approaches to protein spot detection and pattern matching in two-dimensional electrophoresis gel databases. *Electrophoresis* 1999, *20*, 755–765.

[9] Srinark, T., Kambhamettu, C., An image analysis suite for spot detection and spot matching in two-dimensional electrophoresis gels. *Electrophoresis* 2008, *29*, 706–715.

[10] Morris, J. S., Clark, B. N., Gutstein, H. B., Pinnacle: A fast, automatic and accurate method for detecting and quantifying protein spots in two-dimensional gel electrophoresis data. *Bioinformatics* 2008, *24*, 529–536.

[11] Vincens, P., Paris, N., Pujol, J. L., Gaboriaud, C. *et al.*, HERMeS: A second generation approach to the automatic analysis of two-dimensional electrophoresis gels. Part I: Data acquisition. *Electrophoresis* 1986, *7*, 347–356.

[12] Ferry-Dumazet, H., Houel, G., Montalent, P., Moreau, L. *et al.*, PROTICdb: A web-based application to store, track, query, and compare plant proteome data. *Proteomics* 2005, *5*, 2069–2081.