Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography B, 815 (2005) 25–37
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Review

Geometrical distortions in two-dimensional gels: applicable correction methods

T. Aittokallio[a,*], J. Salmi[b], T.A. Nyman[c], O.S. Nevalainen[b]

[a] *Department of Mathematics, University of Turku, FIN-20014 Turku, Finland*
[b] *Department of Information Technology, University of Turku, Turku, Finland*
[c] *Turku Centre for Biotechnology, Turku, Finland*

Received 7 June 2004; accepted 26 July 2004
Available online 15 September 2004

Abstract

Two-dimensional electrophoresis (2-DE) provides a rapid means for separating thousands of proteins from cell and tissue samples in one run. Although this powerful research tool has been enthusiastically applied in many fields of biomedical research, accurate analysis and interpretation of the data have provided many challenges. Several analysis steps are needed to convert the large amount of noisy data obtained with 2-DE into reliable and interpretable biological information. The goals of such analysis steps include accurate protein detection and quantification, as well as the identification of differentially expressed proteins between samples run on different gels. To achieve these goals, systematic errors such as geometric distortions between the gels must be corrected by using computer-assisted methods. A wide range of computer software has been developed, but no general consensus exists as standard for 2-DE data analysis protocol. The choice of analysis approach is an important element depending both on the data and on the goals of the experiment. Therefore, basic understanding of the algorithms behind the software is required for optimal results. This review highlights some of the common themes in 2-DE data analysis, including protein spot detection and geometric image warping using both spot- and pixel-based approaches. Several computational strategies are overviewed and their relative merits and potential pitfalls discussed. Finally, we offer our own personal view of future trends and developments in large-scale proteome research.

Keywords: Geometrical distortions; Spot detection; Image warping

Contents

* Corresponding author. Fax: +358 2 333 6595.
E-mail address: tero.aittokallio@utu.fi (T. Aittokallio).

1570-0232/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2004.07.037

1. Introduction

Proteome is defined as 'the proteins expressed by a genome', and proteomics can be defined as 'the large-scale study of proteins'. Therefore, one of the basic requirements for proteome studies is the need for a separation method that is capable of separating very complex protein mixtures, even many thousands of proteins in one experiment. In addition, the method must allow for protein quantification after separation and recognition of the proteins that change as a function some stimulus. Two-dimensional electrophoresis (2-DE) dates back to 1975 [1], but it is still considered as the best method available to fulfil these requirements. In 2-DE, proteins are separated according to their charge and size, respectively, into distinct spots in a polyacrylamide gel. The location where the protein migrates during separation is characteristic for that specific protein, and the size and intensity of the spot is related to the amount of the protein. In proteome studies, separate 2-DE gel is run from every sample, and the resulting gels are compared to find differentially expressed proteins between samples. After comparison analysis the spots of interest can be cut out from the gel and the corresponding proteins identified by using other techniques, most often by mass spectrometry (MS) and database searches [2].

Visualization of the proteins in 2-DE gels after electrophoretic separation can be done with several different techniques, e.g. by Coomassie Brilliant Blue (CBB) staining, silver staining, or by using fluorescent dyes such as SYPRO Ruby. In cases where the proteins have been in vivo labelled with radioactive isotopes, such as ^{35}S or ^{32}P before 2-DE, also autoradiography detection can be used. The traditional CBB stains almost all the proteins with good quantitative linearity, but the sensitivity of CBB is not very good. Silver staining can detect protein amounts even in the sub-nanogram level (down to 0.2 ng), but for quantification it suffers from low linear range of the stain. Also, due to technical reasons it is difficult to stain gels into the same total intensity. Fluorescent dyes are at present almost as sensitive as silver staining, and they have wide linear dynamic range for quantification. However, these stains are expensive and require special equipment for visualization of the proteins after staining. Radioactive labelling is the most sensitive detection method and also best for quantification, but it requires living cells as starting material, and therefore many other biological sample types cannot be used in combination with this method. See [3,4] for reviews of visualization techniques.

Even though 2-DE has been the true working horse in most proteome studies published so far, it suffers from some drawbacks. The separation method requires a lot of manual work and it is difficult to produce reproducible 2-DE gels even from the same sample. Comparison analysis of digitized 2-DE gels is most often done with specialised software, but it is still not a trivial task as a single gel may consist of many thousands of noisy spots. One of the major obstacles when analyzing 2-DE gels originates from the complex nature of the data [5]. The changes in observed data between different gels originate from both biological variation (corresponding to the true differences between different cell types, tissues or individuals) and technical variation (corresponding to systematic noise from the technique itself). Therefore, the gels may represent spatial variability within them so that the same protein spots may have different location in different gels. The main sources of experimental noise contributing to such distortions are due to the differences in sample composition, casting and polymerization in addition to 2-DE gel running, staining, and scanning. Several confounding factors in 2-DE technique have been identified such as the structure of the polyacrylamide net, the characteristics of the transporting solute, the solvent conditions, and the nature of electric field used [6].

In the present paper, we survey the main concepts behind the software packages for 2-DE gel analysis, with particular focus on the image analysis methods, which aim at adjusting for any systematic geometric distortion inherent in the digitized images. In order to understand these correction methods it is necessary also to have a basic understanding of spot detection that often precede the spatial correction step, see Fig. 1. The objective of spot detection is to find the meaningful real protein spots on individual gels, and examine their quantification characteristics, including density and amount of protein in a spot. In case of comparison studies, a funda-

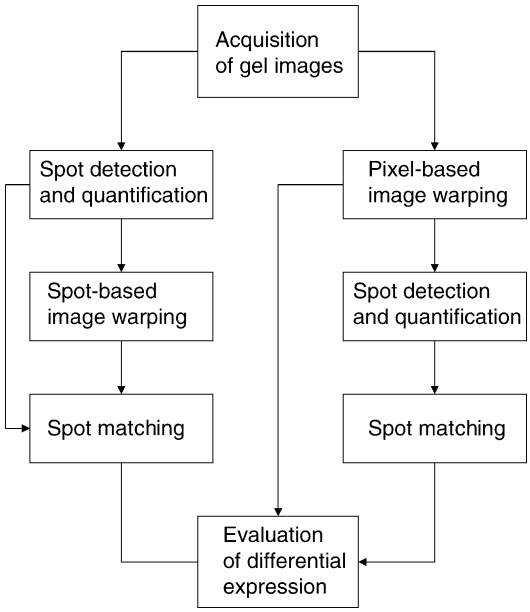


Fig. 1. The two main approaches to the computer-assisted analysis of 2-DE gels. First, spot-based analysis, where the spot detection is the basic step followed by correction of geometric distortions (image warping) and spot matching (the left-trajectory). It is customary to introduce user-defined landmark spots to help image warping, whereas fully automatic spot matching makes the spatial correction implicit in the method. Second, pixel-based analysis performs spatial correction directly from the raw images, and optional spot detection and evaluation of differential expression rely on it (the right trajectory). Spot detection in some form should be employed also in pixel-based methods if the aim is to provide protein quantification information or to pair the proteins between gels automatically rather than using graphical displays.

mental objective is to discover any differential expression between the samples run on a series of 2-DE gels. Gel matching seeks to perfectly align the corresponding proteins in different gels using methods from correspondence analysis. Geometric distortions of the protein patterns may complicate gel matching considerably and image warping is often used to reduce these deformations. Image warping seeks to define a smooth transformation from one gel image to another, where the geometric relationship between gels can be learned manually from user-defined landmark spots or automatically from specific image features. Approaches to image warping (or image registration) can be broadly divided into two categories (Fig. 1). In the customary analysis workflow, spot features are first extracted from the images and the warping process is applied to them, while in the other category, the warping process is applied directly to the intensity profiles of the two images under comparison.

2. Software packages

As the 2-DE experiments result in a wealth of noisy data, efficient utilization of this technique relies on the use of automated image processing techniques. Therefore, the overall success of the 2-DE-based proteome research depends critically on the accuracy and the reliability of the analysis software used to process the data. Moreover, the software has a profound effect on the interpretation of the results obtained and on the amount of user intervention needed during the analysis task. The inherent sources of both biological and technical variation in the experiments pose a great challenge for the analysis algorithms to cope through the multistage process involving steps like contrast enhancement, background subtraction, artefact removal, spot segmentation, expression quantification, landmark pairing, image warping, automatic matching, and finally identification of the differential protein expression between samples under different experimental conditions.

With the risk of oversimplifying the matters one can separate three generations of software for 2-DE gel analysis [7]. The first generation, dated to the second half of seventies and early eighties, used mainframes and minicomputers mostly without a programmable graphical interface. ELSIE [8,9], GELLAB [10,11], TYCHO [12], and LIPS [13] will be mentioned here as examples belonging to this early era The facilitated use of graphical user interfaces such as XWindows and the development of modern operating systems such as Unix enabled the revision and design of second generation analysis systems in the late eighties. These software were often developed in innovative research laboratories and run in general-purpose hardware and software environments. Computer methods such as Elsie-4 [14] (based on Elsie), Melanie [15] (based on Elsie-4), and QUEST [16] are representatives of this generation that will be discussed here. Some of these early software were also turned into commercial products, such as PDQuest (based on QUEST), Kepler (based on TYCHO). Other commercial software available include Phoretix 2D and Progenesis (Nonlinear Dynamics Ltd.).

The third generation of analysis algorithms appeared because of the fast decrease in hardware prices, the availability of efficient scanners and graphical workstations, the dramatic increase of computation power and the availability of biological databases through Internet. These changes have caused major revisions to existing gel analysis software and the appearance of new easy-to-use software that provide sophisticated data visualisation and downstream analysis tools. We will describe in particular some features of Melanie II [7] (based on Melanie), CAROL [17] (fully automatic), Z3 [18] (hybrid pixel/spot-based) and MIR [19] (pixel-based method). These algorithms utilize the advances from image processing, computer vision and machine learning research to allow efficient treating and accurate comparison of 2-DE gel image data with increasing complexity. Other features of the recent analysis packages include possibility to link information at different levels, e.g. management of mass spectrometric and 2-DE data in an integrated way, and the portability to many different levels of hardware platforms, ranging from efficient workstations to PCs. A rather comprehensive list of 2-DE analysis packages on the market today along with their specific features is presented elsewhere [20].

2.1. Geometric distortions

Although the advances in 2-DE technology have significantly improved the reproducibility of 2-DE results, gel-to-gel variability still exists such that the spot patterns in different gels cannot be directly superimposed. Geometric distortions introduce differences in coordinates of two identical points in the reference and distorted gel. Let $(x_i^{\mathrm{R}}, y_i^{\mathrm{R}})$ and $(x_i^{\mathrm{D}}, y_i^{\mathrm{D}})$ represent the x- and y-locations of the corresponding point on the reference gel and distorted gel, respectively, for N positions $i = 1,2, \ldots, N$. Then the distortion vector connecting the two locations can be represented by polar coordinates $(r_i^{\mathrm{RD}}, \theta_i^{\mathrm{RD}})$, where the Euclidean distance between the positions is:

$$r_i^{\mathrm{RD}} = \sqrt{(x_i^{\mathrm{R}} - x_i^{\mathrm{D}})^2 + (y_i^{\mathrm{R}} - y_i^{\mathrm{D}})^2}, \tag{1}$$

and the angle relative to the horizontal axis is:

$$\theta_i^{\mathrm{RD}} = \arctan\left(\frac{y_i^{\mathrm{R}} - y_i^{\mathrm{D}}}{x_i^{\mathrm{R}} - x_i^{\mathrm{D}}}\right). \tag{2}$$

Variation in these two features over all positions $i = 1,2, \ldots, N$ can be used to quantify the complexity of the transformation between the coordinate systems at different locations of the superimposed gels. The average squared error:

$$E_N^{\mathrm{RD}} = \frac{1}{N}\sum_{i=1}^{N}(r_i^{\mathrm{RD}})^2 \tag{3}$$

provides a measure of overall degree of geometric distortion between the gel pair.

Three basic types of geometric distortions were previously characterized: unordered, ordered and translational [21]. Unordered distortions do not show any specific overall pattern in the distribution of distortion vectors but some level of order can be found in local regions. In the ordered pattern, the distortion vectors show the same direction θ_i^{RD} but the length r_i^{RD} is dependent on the position on the gel. Translational patterns are typically caused by gel shifts that produce similar θ_i^{RD} and r_i^{RD} values over the whole gel. Pánek and Vohradský observed that the geometric distortions were independent of electrophoretic conditions and other experimental parameters used in their tests [21]. However, it is quite usual that the distortions observed in 2-DE gels cannot be classified into one of these categories, but the gels represent all types of distortions coextensively, resulting in both global and local geometric distortions. For instance in Fig. 2, the examples show rather large global shifts towards upper left in general, but also severe local differences are observed, where the distortion vectors at different locations of the gel may be quite dissimilar or even opposite. Therefore, the geometric relationship between the gel pair can be described accurately within a local neighbourhood only.

Given the problems in the reproducibility of the 2-DE data, comparative analysis can be a very tedious task even with the modern 2-DE software packages. In order to make the problems caused by geometric distortion to the gel comparisons more concrete, we shortly cite the research conducted by Voss and Haberl [22]. The authors studied the efficiency of spot matching using a set of 49 gels produced from mononuclear cells extracted from human blood. They found that only about 90–95% of all spots could be matched in Melanie II software for gels that were produced from the same sample and run in parallel. Overall matching efficiency dropped to 8.9% when counting all the spots that could be matched in at least 40 gels. When selecting one gel as a reference gel and matching all the other gels against this gel, they achieved a pair-wise matching efficiency of 89% despite manual landmarking. The matching efficiency could not be improved by increasing the number of landmarks used for the alignment. Hence, it becomes very difficult to perform gel comparisons with multiple gels as only few spots can be matched in all gels. Voss and Haberl attributed these problems to severe geometric distortions both in local and global scale between spot patterns and to the poor spot detection due to silver staining [22].

2.2. Software comparisons

Since the commercial packages for 2-DE data analysis are nowadays closed source and they are based only partly to the original academic developments without detailed implementation available, evaluation of software products must be based on a set of tests that imitate the common practical problems encountered over the data analysis pipeline. Results obtained from such evaluations can be assessed directly through expert scorer or indirectly through replicate experiments. For instance, spot detection capability can be tested by comparing automatically segmented spots to manually counted "real" protein spots, quantification by analysing a dilution series or artificial images, and spot matching by aligning distorted gels with the original ones. However, as the absolute truth behind the tests is typically unknown and the results highly depend on the data, the effectiveness of a package cannot be evaluated reliably without comparing it to other methods using the same data. It is therefore recommended that the investigators do not make strong claims regarding the significance of new algorithms without comparing them first to more standard methods.

An early study by Myrick et al. compared Visage 2000 and Gellab-II analysis systems within a set of 29 silver-stained 2-DE gels from the study of urinary proteins [23]. They evaluated the software with respect to spot detection, quantification, and matching individually, whereas rigorous inter-system comparison was presented for quantification of nine selected spots only. Mahon et al. [24] evaluated the reproducibility in quantification of the Phoretix 2D software with CBB staining using multiple scans of the same gel. However, no comparison to other software or staining was presented, reducing the significance of these results for software selection. A more comprehensive comparison of three analysis systems, Progenesis, Z3 and PDQuest, in spot detection and quantification, was performed by Nishihara and Champion [25]. They used 2-DE gels stained with SYPRO Ruby to analyze *Escherichia coli* proteins in a design where a cell extract was serially diluted to seven different protein levels spanning a 1000-fold range. This titration was repeated four times and the 28 samples were loaded onto separate gels. The three programs produced similar numbers of spots among the replicates. The coefficient of variation (CV) in their spot detection reproducibility ranged from 4 to 11%, with the exception of the lowest protein load (0.5 μg), where more proteins were around the detection threshold. The reproducibility in spot quantification as evaluated within a set of 20 selected proteins was also comparable across the three programs, with CVs ranging from 3 to 33%, where again the higher values were for proteins of lower abundance. Linearity of dilution series was demonstrated with only single known protein in PDQuest and Progenesis.

Raman et al. [20] compared a spot-based method Melanie (version 3.0) and a pixel-based method Z3. In spot detection, they used as a test material two 2-DE gels provided by the software companies, where the spots were counted manually in order to assess the matching performance reliably. A set of 12 synthetic gels with Gaussian-shaped spots of known volume was used to test the spot quantification. The former gels were also distorted by using 'height decrease' and 'centre pull' distortions in various degrees to produce a set of nine reference-distortion pairs for testing spot matching. In spot detection, Z3 performed better than Melanie. The results in spot matching depended on the type of distortion. For geometric distortions, Z3 gave better results, whereas in non-geometric distortions, the packages performed comparably. Melanie was better in spot quantification. They

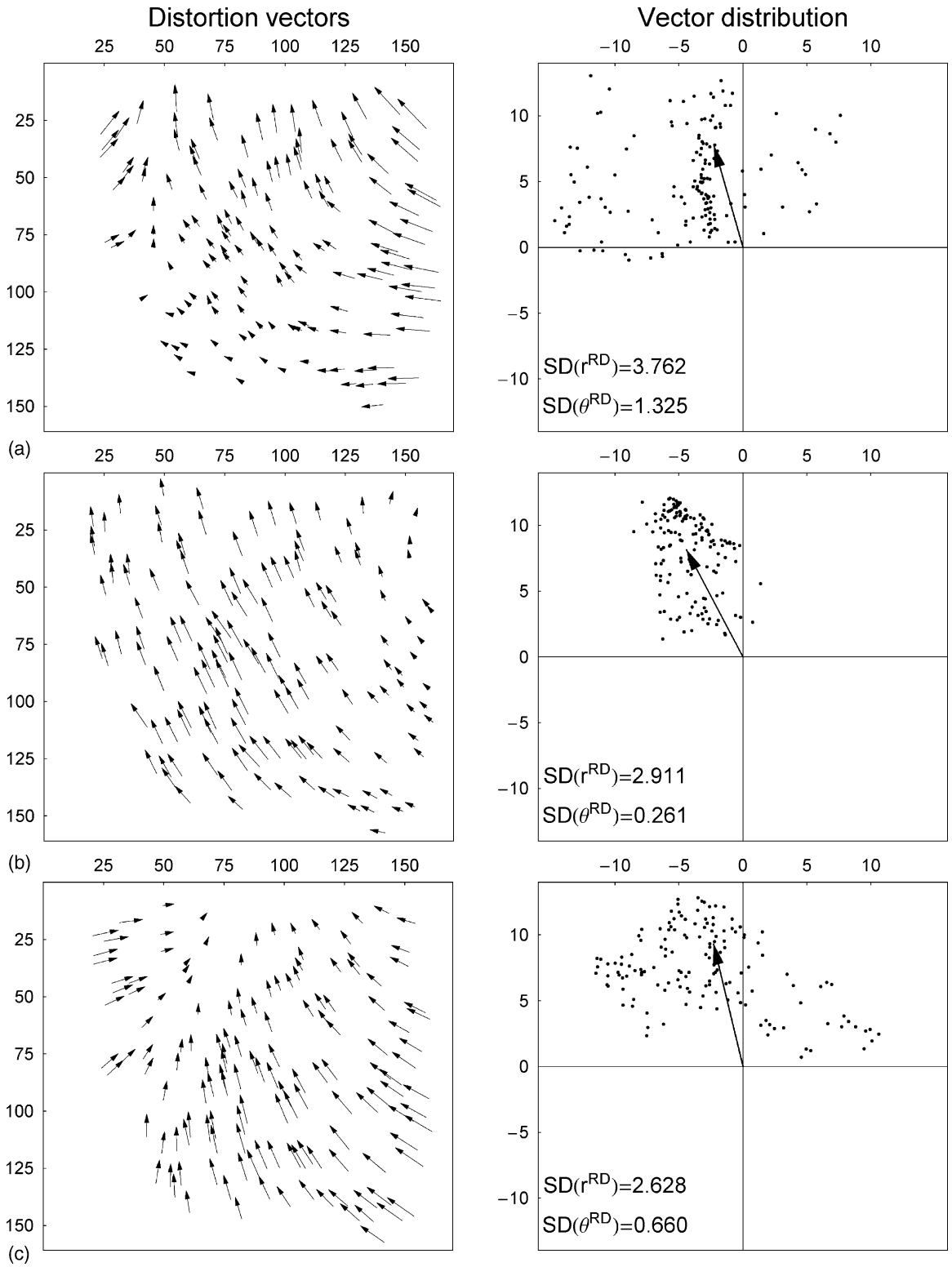


Fig. 2. Three examples of geometrical distortion between gel pairs, reproduced from [34]. Left panel: distortion vector plot, where the start of the arrow shows the location of the spot on the reference gel (R) and the end on the distorted gel (D). There are N = 144 manually matched points in the figures. Right panel: vector distribution plot, where the starting point of the vectors is shifted to the origin and the endpoints are marked as dots. The arrow indicates the global mean distortion vector. Standard deviation (S.D.) both for the vector length r^{RD} and its angle θ^{RD} is marked within the pictures (see Eqs. (1) and (2)). The overall level of geometric distortion varies markedly across the three example pairs, as measured using average squared error (Eq. (3)): (a) E_N^{RD} = 78.205, (b) E_N^{RD} = 94.123, and (c) E_N^{RD} = 99.000. The unit in the figures is millimetres. Reproduced with permission from [34].

also published their test material to build a standard test benchmark for other comparisons studies. Rosengren et al. [26] utilized the same test procedure and material as Raman et al. in a comparison between PDQuest (version 7.0.1) and Progenesis (version 2002.1). In addition, they used three real-life gel sets including repetitions from the same sample, samples after different treatments, and artificially distorted gels. Both of these programs perform an automated spatial correction without user-specified landmarks. Although there was no significant difference between PDQuest and Progenesis, it was observed that both packages were sensitive to the adjustable parameters with respect to the tendency of finding true positive and false positive spots. When comparing to the results of Raman et al. [20], Z3 was found to be the best in spot matching. Spot quantification was distinctly more accurate with Progenesis and PDQuest than with Melanie and especially Z3.

3. Spot detection

A principal goal of 2-DE technique is to discern individual protein spots from the gel images, followed by biological questions concerning whether the same protein occurs in the gels of interest. Therefore, it is quite natural that the fundamental step in the traditional workflow of 2-DE data analysis is the spot detection process and the subsequent steps are based on the gel-specific lists of spot coordinates (Fig. 1). By spot detection it is possible to reduce the amount of data from millions of image pixels to some thousands of spot-features. This was considered crucial when computational power was not as high as today. Recently, there have been a growing number of studies suggesting that the traditional workflow may not be the optimal one. Instead, the gel comparison process can be applied directly to the raw data acquired from the gels (see Section 4.2). Nevertheless, whether a spot-based or pixel-based method is used, some sort of spot detection from the gels must be used in order to enable protein-specific quantification information.

There are three major problems in accurate spot detection. The first problem is concerned with the technical noise originating from the image acquisition process. In Melanie system, for example, smoothing with local kernel can be used to reduce high frequency noise inherent in the acquired images and histogram equalization and contrast enhancement are options to improve the difference between spots and background [27]. After noise removal, background subtraction is applied to eliminate meaningless changes in the gel background intensity level. The background varies in different parts of the gel causing disappearance of weak contrast spots or merge of nearby saturated spots. Melanie estimates the background by fitting the pixels located outside the spot regions with a third-order polynomial function [27]. More advanced technique originates from mathematical morphology, where the background variability is estimated by sliding a structuring element, so-called ‘rolling ball’, under the image [6]. The pre-processing methods are naturally strongly dependent on their adjustable parameter values.

The second problem in reliable detection of spots is due to the difference in the mobility in the two orthogonal electrophoretic dimensions that usually results in spots that do not appear as ideal rounded forms. Although it has been asserted that the principal form of the density distribution of spots is the two-dimensional Gaussian type, in practice, spots can be oblonged or may even have long tails in both directions. Therefore, more complex spot models should be exploited in spot segmentation to improve the sensitivity. The third problem is caused by the existence of true overlapping or touching spots [28]. In the case of complex spot patterns, simple thresholding rarely works in practice [10]. This will impose major challenges to the matching step as complex spot region may be interpreted as a single spot in one gel whereas in another it may be separated into several different partially overlapping spots. For those cases it may be advantageous to have several alternative interpretations of complex spot patterns as in CAROL system [29].

3.1. Spot segmentation

The aim of the segmentation process is to define the location, boundary and intensity for each spot present in a gel. Early methods often applied nonparametric models for spot centre detection, based on Laplacian transform and second derivatives, and parametric models for spot modelling, such as Gaussian functions and 2D polynomials. The spot segmentation method of the GELLAB system [10], for example, relies on the properties of derivatives on smooth surfaces and the interpretation of the intensity table as a three-dimensional surface. The idea is to recognise peaks of the surface and segment them according to the changes of the inclination when moving down-hill, see Fig. 3. The second derivative I of the intensity function I in a chosen direction has typically three peaks: one large corresponding to the spot centre and two smaller ones at the foot of the peak. Boundary of the spot is determined by the points of the smaller maxima. On the detailed technical level, the spot segmentation still contains a number of steps aiming at separation of touching or overlapping spots, removing concavities and filling corners of spots. The basic idea in separating spots is to perform the segmentation as a two pass process where the first pass restricts the boundaries to the points where I'' becomes increasing, marking these pixels belonging to spot core regions. The second pass expands the cores to regions where I magnitude function has its second local maxima (Fig. 3).

In the Elsie-4 system [14], the spots are detected by thresholding the so-called peakedness values at each point instead of the original intensities. If $V(x, y)$ stands for a smooth surface fitted over the raw intensity observations I, one can define the peakedness of V as the negative of the Laplacian:

$$P = -\left(\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2}\right). \quad (4)$$

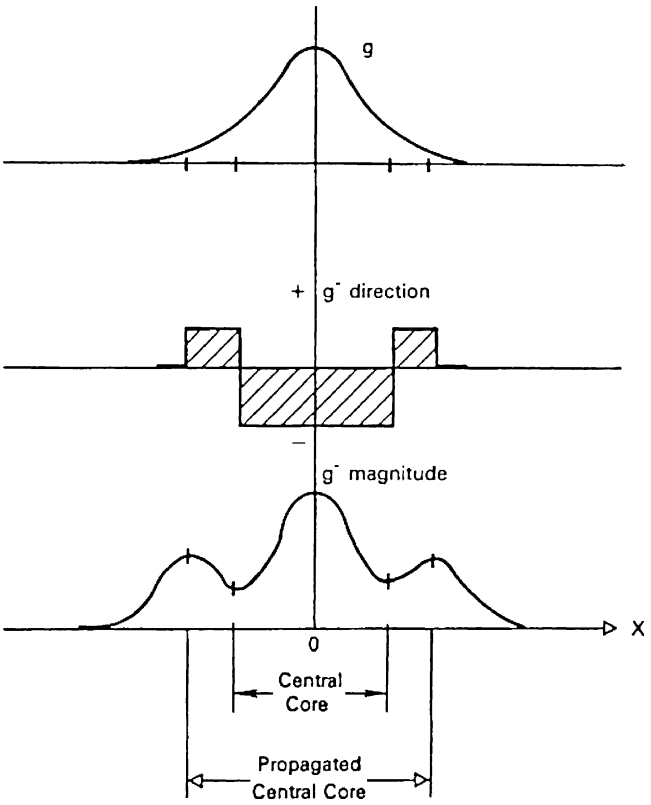


Fig. 3. The use of the second derivative I'' for distinguishing the spot region from the intensity surface of an image I. Reproduced with permission from [10].

By stepwise thresholding the $P(x, y)$ values one can effectively separate spots that touch each other. Melanie II [27] performs the spot detection by a modification of the Laplacian thresholding and second derivatives. A point (x, y) with intensity $I(x, y)$ and Laplacian $P(x, y)$ belongs to a spot if

$$\min\left(\frac{\partial^2}{\partial x^2}I(x, y) - C_x, \frac{\partial^2}{\partial y^2}I(x, y) - C_y\right) > 0 \quad \text{and}$$
$$-P(x, y) - L \geq 0. \tag{5}$$

Here, positive constants L, C_x and C_y are thresholds for the Laplacian image (Eq. (4)), and for the second derivative along the x- and y-axis, respectively. For saturated pixels, the condition in Eq. (5) is modified accordingly [27]. A disadvantage of these edge detection algorithms is the possibility of false detection of artefacts if their boundaries have similar characteristics to those of real protein spots. Subsequent manual removal and editing of the spots can be laborious and it introduces also undesirable subjectivity to the analysis.

Recently, other methods have been suggested that overcome some limitations of simple edge detection. A popular method for spot segmentation in pattern recognition research is the watershed transform (WST) [30]. Imagine drilling holes in each local minimum of the landscape and immersing it into a lake, see Fig. 4. The resulting 'watersheds' define the optimal contours of objects under investigation. CAROL system uses the WST on the surface of the first derivative of

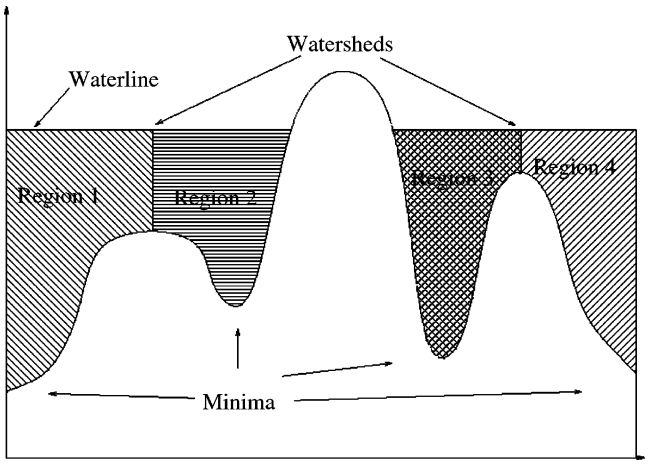


Fig. 4. Operation of the watershed transformation. Modified with permission from [17].

the 2-DE gel pixel intensities I [17]. The assumption is that valleys of the gradient image correspond to the requested regions whereas the ridges define the boundaries of a region. The WST results in well-located closed contours, but also in tendency for strong over-segmentation due to experimental noise creating false minima. To exclude the false regions among the candidate ones, CAROL considers only regions R with convex curvature, thus satisfying.

$$\sum_{(x,y)\in R} I''(x, y) > 0. \tag{6}$$

The merging process of partial spot regions continues by requiring that each valid spot should have an approximately elliptical shape as assessed with χ^2-test [17]. Such iterative approach is feasible these days because of increased computation power available.

3.2. Spot quantification

After defining the boundaries of spots, the next step is to determine their quantification features such as area, volume and density. Spot quantification can be achieved using at least two approaches: parametric and non-parametric methods. Perhaps the most popular parametric model used for 2-DE spot quantification is the two-dimensional Gaussian function. In QUEST system [16], for instance, Gaussian curve is fitted to each spot optimizing the location (x, y), amplitude A, and deviation along the axes (σ_x,σ_y). A special care must be paid here to the modelling of overlapping spots, which is done automatically by iterating the surface estimation and subtracting Gaussians of overlapping spots. The estimated curve can be integrated to yield the parametric volume of the spot

$$\text{PV} = \pi A\sigma_x\sigma_y. \tag{7}$$

In addition, QUEST includes also a possibility of manual intervention to combine a mixture of Gaussian curves to present

a single spot, whereupon the total volume is calculated by summing Eq. (7) over all the components.

Non-parametric spot quantification considers the intensities I inside the segmented boundaries of each spot region R. Melanie II system [27] allows the calculation of several non-parametric quantification features in addition to Gaussian fitting. Examples of these include optical density:

$$\mathrm{OD} = \max_{(x,y)\in R} I(x, y) \tag{8}$$

and non-parametric volume:

$$\mathrm{NV} = \sum_{(x,y)\in R} I(x, y). \tag{9}$$

Values of PV, OD, and NV are typically normalized by the overall value over the gel when quantifying individual protein expression. It has been shown that the relative volumes provide more accurate estimates of the true protein amount [31]. However, none of the features consider background stain levels and they all have a limited range of linearity. To overcome these problems Dutt and Lee [32] introduced a composite of the area and volume features, so-called scaled volume (SV). Calculation of the SV values, however, requires manual editing for selecting spots that are not of interest (e.g. technical artefacts).

4. Image warping

In a typical research setting, the principal goal of 2-DE analysis is the identification of differentially expressed proteins between samples run on different gels. An important prerequisite for efficient gel matching is therefore the image warping step, where the geometric relationship between the gels is modelled through a transformation which maps all positions in one image to positions in a second image. This problem arises in many image analysis problems, whether one seeks to remove geometric distortions, to register an image with a reference, or to align several images. See [33,34] for general surveys on image warping and registration techniques. The choice of the warping function is a compromise between a smooth transformation and one which achieves a good match. While the latter aims at maximizing the sensitivity of the matching process, the first one can be used to control for the specificity of the matches; function with too many adjustable parameters can be easily overfitted to match a limited set of corresponding locations only.

Warping of the 2-DE gel images can be carried out by two different approaches: spot-based and pixel-based methods. Spot-based methods start with the given list of detected spots and the actual warping then considers the spots as individual points and the task is to find a transformation that maps the gels in question to resemble each other. This process often involves the use of so-called landmark points to guide the search of a good transformation function. These are point pairs that the user manually determines to correspond in the different gels. In the semi-automatic systems, landmarks are used as anchors for the distortion correction process, and they are frequently used as starting point for matching as well. Landmarks are often among the largest spots, and a general recommendation is to select them so that they cover the gel evenly [11]. One can also choose those landmarks spots that are supposed to facilitate the recognition of corresponding spot patterns in the reference and test gels. In some analysis systems, the selection of landmarks is automatized, thus removing some of the manual labour.

Pixel-based methods perform the image warping directly on the raw data, by considering the image as a surface formed by the pixel intensities, rather than indirectly by using the detected spot lists. The images can then be displayed in an overlaid fashion using a colour scheme to highlight the intrinsic structural differences between the gel pairs [19]. The motivation for such direct methods comes from the observation that spot-based methods use only a fraction of the available information from the images when going through an intermediate and often noisy step of spot detection. When basing the warping on the raw pixel values, numerous additional features, such as spot shape and intensity spread, that are otherwise lost in the spot detection, are available for correcting geometric distortions. This approach is computationally rather demanding as the intention is a shift in complexity from spot detection to the image warping phase. It was not until recently when the computational power became cheap enough for allowing direct registration methods [18].

4.1. Spot-based warping

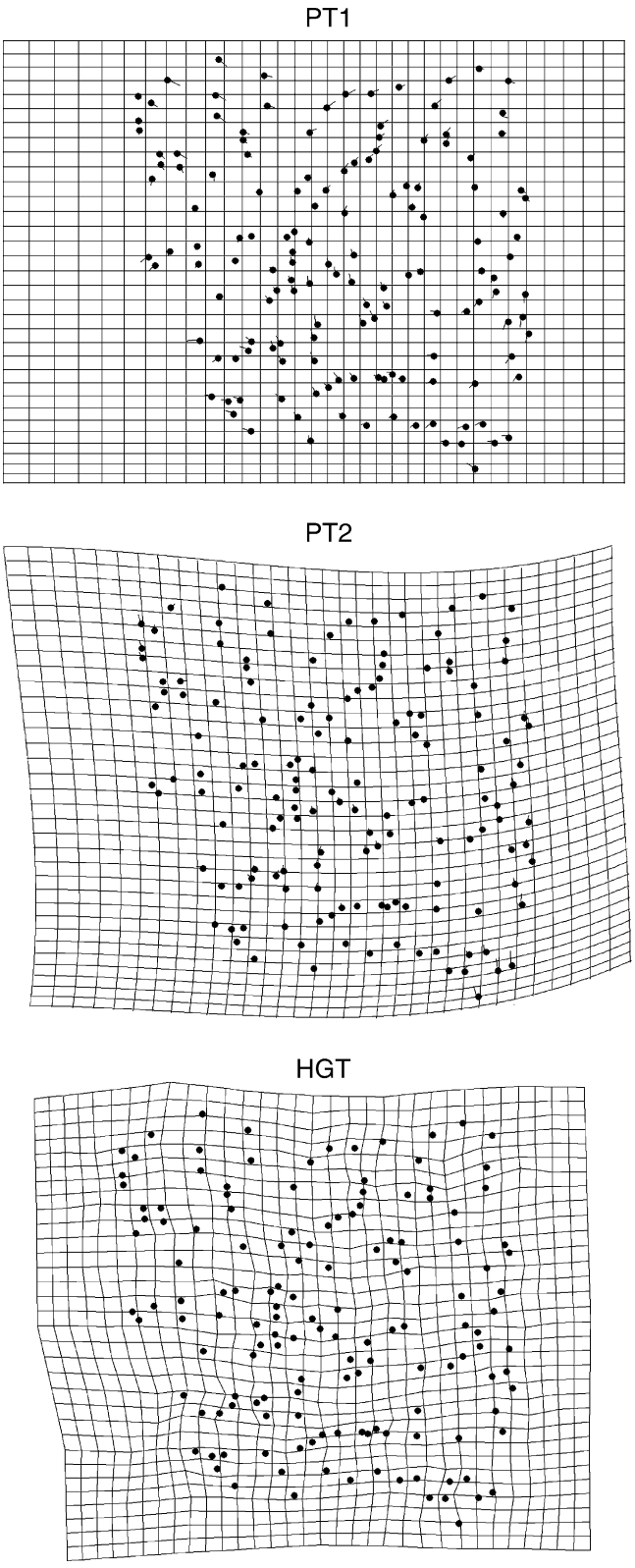
Traditional methods for spot-based warping often use polynomial functions to align the spots in two gels, whereas nowadays it is know that such global approach is incapable of modelling the complex geometric distortion inherent in 2-DE gels. In polynomial warping, we assume that the geometric relationship between the two coordinate systems, reference (R) and distorted (D), can be modelled by a linear combination of given M functions:

$$x^{\mathrm{R}} = \sum_{j=1}^{M} a_j f_j(x^{\mathrm{D}}, y^{\mathrm{D}}) \quad \text{and} \quad y^{\mathrm{R}} = \sum_{j=1}^{M} b_j f_j(x^{\mathrm{D}}, y^{\mathrm{D}}). \tag{10}$$

Given N known spot pairs (x_i^R, y_i^R) and $(x_i^{\mathrm{D}}, y_i^{\mathrm{D}})$, obtained either from manual landmarking or automatic spot pairing, the coefficients a_j and b_j can be determined by minimizing the error function in Eq. (3), provided that $N \geq M$ [35]. The basis functions $\{f_1, f_2, \ldots, f_m\}$ can be widely non-linear functions on both x and y, but typically low-order monomial functions $f_j(x, y) = x^{m_j} y^{n_j}$ are used to ensure the smoothness of the warp. For polynomial functions of order n, there are $M = (n + 1)(n + 2)/2$ coefficients to be determined for both dimensions. The early version of ELSIE [8], for example, used linear warping functions, that is, $n = 1$, and therefore three or more reference points (landmarks) must be provided. Pánek and Vohradský experimented with various number of n and

demonstrated that the best polynomial degree was $n = 3$ in their test material [21].

Because in multiple gel comparisons the polynomial coefficients are determined several times, it is sometimes useful to use simply monomials of one variable only, i.e. $f_j(x) = x^{m_j}$ and $f_j(y) = y^{n_j}$. A polynomial of the order n can now be estimated by using at least $M = n + 1$ landmarks. This simplication is applied, e.g. in Melanie II [27]. However, it is evident that there are cases where such global transformation with low-order polynomials cannot correct the local geometric distortions, see Fig. 5. As poor warping results increase the complexity of the matching phase, several improvements to this global scheme have been developed. In TYCHO system by Anderson et al. [12], the gel warping is performed by a series of local deformations whose effect decline exponentially with respect to the distance to the centre point. In QUEST system by Garrels [16], there are three different ways of performing the polynomial transformations. For densely-populated neighbourhoods, the warping is simply the average shift between the two coordinate systems ($n = 0$). For sparser areas, both translational and scaling factors are computed ($n = 1$). In distorted regions, second order correction factors are also generated ($n = 2$). Such local transformations should facilitate the correction of local geometric distortions, provided that the correction model is flexible enough for all the deformations observed in the 2-DE gel pairs.



To satisfy the need of more flexible distortion correction, several authors have tested new methods to warp 2-DE gels, originating mainly from image processing research. Horgan et al. [36] compared thin plate spline (TPS) transformation to the global linear polynomial warping. TPS incorporates an extra function that allows nonlinear bending of the gel coordinates [37]. As expected, TPS transformation led to better matching results, perhaps due to the fact that TPS provides exact matching of the training spots used. Salmi et al. [35] designed an approach to multiresolution, piecewise bilinear mapping, using so-called hierarchical grid transformation (HGT). The idea of HGT is to iteratively subdivide the initial quadrilateral superimposed on the distorted gel into grids of smaller convex quadrilaterals, see Fig. 6. The corners of the grids are optimized at each iteration step using random descent method. The hierarchical processing provides a single model for both global distortions (modelled during the early steps) and local distortions (modelled at the end), provided a sufficient number and distribution of corresponding spots is available. A disadvantage of HGT is the large number of grid points to be optimized and possibility to overfit to the training data. When comparing HGT to the global polynomial transformations of order three using cross-validation (Fig. 5), however, they observed superior warping efficiency with HGT already at small training set sizes ($N = 15$).

Fig. 5. Transformation of a distorted gel grid using the $N = 144$ corresponding spots of Fig. 2c. The reference location of the distortion vectors is marked by dot and the corrected location is at the other end. The gel pair presents severe local distortions that are poorly modelled by global, polynomial warping functions. PT1: warping with third-order polynomials of one variable along both axes. PT2: warping with third-order polynomials of two variables along both axes. As a comparison, more accurate warping can be achieved with hierarchical grid transformation (HGT), which uses piecewise bilinear warping functions. The relative errors of three warping methods were 5.3%, 1.6%, and 0.001%, respectively. Reproduced with permission from [34].

4.2. Pixel-based warping

Pixel-based warping is typically obtained by maximizing the correlation between two intensity surfaces (I_1, I_2). The pixelwise correlation coefficient between two digital images corr is defined as:

$$\mathrm{corr}(I_1, I_2) = \frac{\mathrm{cov}(I_1, I_2)}{\sqrt{\mathrm{cov}(I_1, I_1)\mathrm{cov}(I_2, I_2)}}, \quad (11)$$

where cov is the covariance between two images

$$\mathrm{cov}(I_1, I_2) = \frac{1}{|D|} \sum_{(x,y) \in D} (I_1(x, y) - \bar{I}_1)(I_2(x, y) - \bar{I}_2). \quad (12)$$

Here, D is the domain of the points to be considered for the registration process. An advantage of using corr as the measure for the similarity is that it is invariant to the changes both in the means $\bar{I}_1$ and $\bar{I}_2$ and in the variances $\mathrm{cov}(I_1, I_1)$ and $\mathrm{cov}(I_2, I_2)$ over the domain D in the two images. The first pixel-based warping procedure for 2-DE gels was presented already in 1992 by Conradsen and Pedersen [38]. They introduced a multiresolution approach with resampling by cubic convolution in order to remove global distortions at lower resolutions and local distortions at higher resolutions. The disparity between the gels is estimated by minimizing sum of squared difference, which basically is equivalent to cross-correlation technique. Starting at 64×64 resolution the images are then repeatedly warped with increasing resolution until 512×512, which corresponds to relative deformation of 5% [6]. However, no smoothness constraints were considered.

Z3 [18] was the first commercial analysis package to perform the raw data based image registration. In fact, Z3 utilizes also spot detection as the registration begins by determining a sequence of covering rectangles each containing a small cluster of spots. This is done by first segmenting the spots in the gels with a simple algorithm such as the second derivative thresholding. The spot segments along with additional features are stored and a score taking into consideration the number of spots, their contrast and area is attached to each rectangle. This score is used for ruling out rectangles strongly overlapping with higher score rectangles. A high scoring rectangle at the centre is the first to be considered for pixel-based, local shift-warping, and the others are considered in order of increasing distance from this seed rectangle, see Fig. 7. The collection of all shift vectors from the regions processed so far is used to generate a global transformation. As more vectors are added the original identity transformation becomes first translational (shift), then linear (rotation and scaling), and finally a Delaunay transformation (piecewise bilinear map-

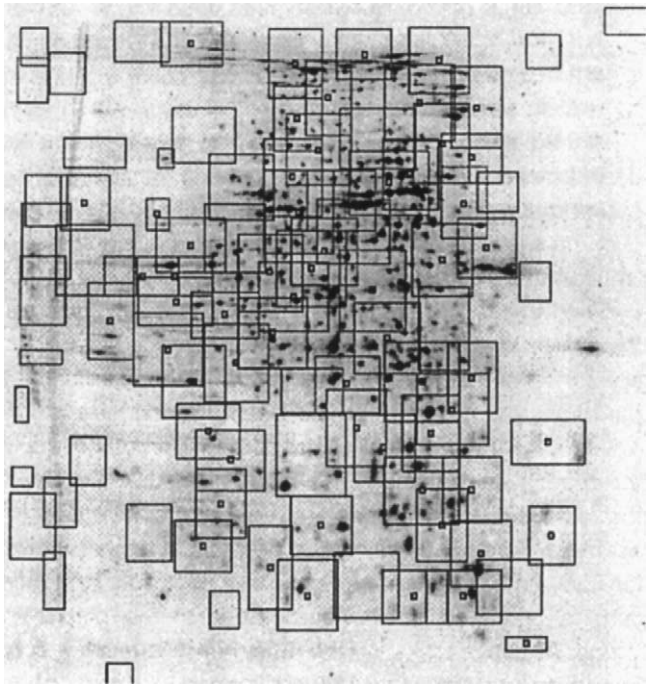


Fig. 7. Rectangle covering of a gel image in Z3. Reproduced with permission from [18].

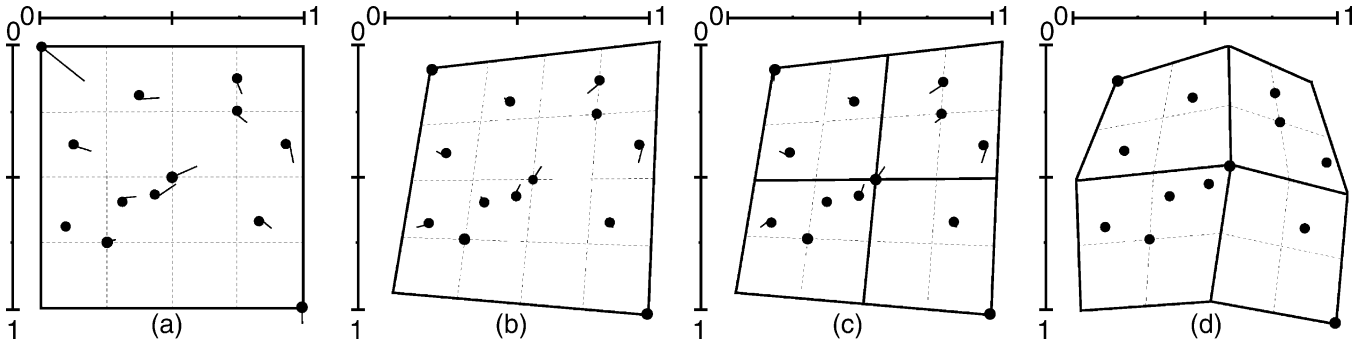


Fig. 6. Schematic illustration of the piecewise bilinear mapping of the spots from the distorted coordinate system to the reference system using the hierarchical grid transformation (HGT). Dots denote the mapped locations of the distorted points, and the distortion vectors are headed towards the corresponding reference landmark location. (a) The most extreme landmark locations define the initial quadrilateral that is superimposed on the distorted gel system. (b) The four corner points of the initial grid are moved to align the corresponding landmark locations by minimizing the average squared distance. (c) The initial quadrilateral is subdivided into four identical quadrilaterals by connecting the centre points of the sides of the initial quadrilateral (dashed lines). (d) The nine corner points of the resulting higher-level grid are optimized based on landmarks within the quadrilaterals. The number of grid points grows in powers of two, until one of the stopping criteria is satisfied. Reproduced with permission from [35].

ping). The global shift, rotation, and scaling factors must be within the pre-defined limits, or otherwise landsmarks are necessary. Once an iteration of the rectangles does not change the local shift list, the process is ready and the final transformation will be used for mapping the individual pixels.

Veeser et al. [19] presented an alternative way of performing raw-data registration of gel images. The so-called multiresolution image registration (MIR) is a direct registration algorithm without any need for spot detection or landmarking. Similar to HGT [35], also their method uses a hierarchical approach which implements the coarse-to-fine warping paradigm to a piecewise bilinear mapping. Instead of random descent, they utilize the gradient of the correlation function (Eq. (11)) and efficient optimization algorithm for determining the best mapping at each resolution. MIR tries to avoid getting stuck into local maxima by using gradually increasing resolutions when refining the grid transforms. Although MIR shows a significant improvement in terms of better registration performance when compared to Z3, the results were not satisfactory for 19% of the 2-DE gel pairs analyzed. Moreover, there were gel pairs that could not be matched due to severe local distortion. Recently, a rather similar two-step approach to pixelwise warping was proposed by Gustafsson et al. [39]. Firstly, the effect of current leakage across the gel sides is described with a physicochemical model and the individual images are spatially corrected to remove this effect. Secondly, the corrected image pairs are automatically transformed using a MIR-type algorithm, where the compromise between achieving a good match and introducing smooth deformations is formulated as to optimize a penalized likelihood criterion [40].

5. Discussion

Image analysis of digitized gels is a critical step for successful 2-DE-based proteome research. At present, image analysis is still one of the bottlenecks in 2-DE studies, as it is time consuming and requires manual assistance with the software programs available. We presented in this paper a general review of the basic techniques in image analysis of 2-DE gels, where the main foci were the spatial correction techniques. Spot segmentation and quantification techniques were also mentioned as they conventionally form an integral part of the 2-DE data analysis task. Two main categories can be recognised in this task: conventional spot-based analysis and advanced pixel-based analysis. It seems that both approaches have still notable practical problems. Even if successful analysis is possible with each technique for good quality gels, more differences between the methods appear when there exist more technical variation (e.g. increased background noise) or biological variation (e.g. different biological conditions). An appealing aspect of the spot-based methods is that they summarize the 2-DE gels as lists of spot features comprising spot location, area, and volume. The lists originating from different gels can then be used in the spot matching step. There is a wide range of automatic matching systems that make the spatial correction implicit in the method by using different correspondence techniques, e.g. based on Gabriel graphs [41] (LIPS system), regular grammars [42] (HERMeS), or Delaunay triangulation [43] (CAROL). Spot alignment provides a matched list of spot features that can be subjected to further classification analysis using e.g. principal components, neural networks, or multidimensional scaling [44–46]. An inherent disadvantage of spot-based methods is that precise segmentation of all spots is computationally intensive as well as error prone, which complicates the spot matching process and can lead to false recognition results.

Perhaps the approach with the greatest potential today includes the pixel-based methods, because they operate directly on the original intensity data, and therefore will not lose any information valuable for correcting distortions and matching gels. This will also permit a more realistic modelling of the gel formation process and systematic errors such as current leakage and local distortions. However, the idea of using the additional geometric information present in the intensity distribution can be a potential pitfall as well, in case the extra information is strongly misleading [19]. Since there are currently only a few methods employing the pixel-based approach and they are under constant development, the strengths and weaknesses of this approach have not yet been sufficiently elucidated. Spot-based correspondence and registration are also subject to active research in pattern recognition research [47–49], so it is likely that more sophisticated methods will be introduced in near future that apply the both approaches. Consequently, well-defined test experiments will be needed to fully compare the different methods. A problem is that there exists no standard protocol for comparing 2-DE analysis packages, but different researchers have used their own non-standardized gel sets and varying comparison methods. A recent improvement towards a ‘benchmark’ of comparison studies is due to Raman et al. [20], despite some methodological limitations. For instance, certain extent of subjectivity will always remain when using an expert for counting the number of spots or scoring the matches provided by the algorithms [26]. By using several experts and double blind testing, the inter-scorer variability could be estimated to assess the real significance of the comparison statistics. Furthermore, although in computer-generated artificial gels, the exact number, quantity, and pairing of the “proteins” is completely known, these synthetic gels are currently far from representing the real world situation. Recently, Rogers et al. [50,51] introduced a more realistic artificial image model for objective evaluation of spot detection.

Development of automatic or semi-automatic strategies for minimizing the geometric distortion associated with 2-DE data presents one of the greatest challenges in the 2-DE data analysis. An adoption of hierarchical or multiresolution approach to image warping scheme seems as the best option because it permits the removal of geometric distortions both at local and global scale. At the moment, fully automatic pixel-based warping with multiresolution representa-

tion can successfully correct for high-order and nonlinear geometric distortion provided that the local deformations are smooth enough [19]. On the other hand, hierarchical spot-based warping with manually defined landmarks is able to correct also radical local distortions provided the user has introduced some landmarks in the critical gel regions [35]. As a possible future development, it would be interesting to combine these two approaches to provide more accurate representation of the geometric relationships between 2-DE gels. A problem here is the labour and time needed for defining even a small set of landmarks, especially when comparing a large number of gel pairs. Although some package such as PDQuest and Progenesis allow nowadays also computer-based definition of landmark points, the landmarking process whether manual or automatic is complicated and prone to errors. In some cases, it is difficult to distinguish the true biological variation from experimental variation, e.g. to determine whether the change in the location of a spot should be attributed to an experimental distortion of the gel or to a biological modification of the protein. Experimental variation influences the protein abundance as well, such that the same amount of the same protein can have different spot intensities on different gels.

An extension to the standard 2-DE technique is the difference gel electrophoresis (DIGE) [52], which aims at minimising gel-to-gel variation in comparison studies. In DIGE, two protein samples are derivatized with two different fluorophores, mixed, and run on a single 2-DE gel. Proteins are detected using two different excitation/emission filters generating two different images. These images can then be directly overlaid, and relative quantification between samples performed. At present, DIGE suffers from certain drawbacks. First, the labelling reagents are expensive and scanning of the images requires special instrumentation. In addition, only 1–2% of the total samples are labelled, and the unlabelled proteins migrate into slightly different location in the 2-DE gel. Therefore, the resulting 2-DE gel needs to be stained with some other dye to visualise proteins that will be selected, e.g. for mass spectrometric identification. Moreover, the sensivity of DIGE is worse than, e.g. the sensitivity of silver staining, and the less abundant proteins in the gels usually remain undetected. The quantification results obtained with DIGE can be improved by incorporating a pooled 'standard' sample labelled with a third dye, which is used to normalize the protein abundance measurements across the experiments [53]. If a similar approach to spatial correction for traditional 2-DE gel analysis could be developed, it would dramatically improve not only the gel matching process, but also the management and analysis of 2-DE gels in general, in terms of helping to create an accurate and universal standard for representing, sharing, and intergrating 2-DE gel images and associated proteome data [54].

Proteome studies in general can be considered to operate on three different levels: expression proteomics addresses question like when, where, and how much proteins are expressed, modification-specific proteomics characterizes proteins' post-translational and other modifications, and 'interactomics' studies protein-protein interactions. So far most proteome studies have focused on expression proteomics since there are methods readily available, whereas with the latter two techniques are still mostly under development. However, it is clear that in the future the proteome-wide studies of protein modifications and interactions will be extremely important to complement the results obtained from expression proteomics, and to give a more complete view of the biological system under analysis. In expression proteomics, there are two different strategies available. The traditional way is to use 2-DE gels for protein separation and quantification, followed by mass spectrometry for protein identification. A newer way is bypassing gels and using MS both for protein quantification and identification. One of the most popular MS-based techniques at present is the isotope coded affinity tag (ICAT) technique [55], where proteins from one sample are labelled with light and proteins from another sample with heavy ICAT-label. After labelling, the samples are mixed, digested into peptides, and the resulting peptides are separated using multidimensional chromatography before quantification and identification by mass spectrometric methods. The MS-based methods can be almost completely automated, whereas running 2-DE gels still requires a lot of manual work. In addition, 2-DE gels show systematic bias against some protein classes, including very big and hydrophobic proteins like membrane proteins, very small proteins and proteins' with extreme p*I*'s, so that they are underrepresented or absent in the gels [56]. It has been shown that MS-based methods are suitable also for these protein classes [57].

Despite the above mentioned pitfalls, 2-DE is still the method of choice for separating very complex protein mixtures. In addition, 2-DE gels are capable of separating protein isoforms into distinct spots, whereas in MS-based methods the data from different isoforms is usually lost. Therefore, it is clear that in the future 2-DE and MS-based methods will be used as complementary tools in proteome studies.

References

[1] P.H. O'Farrell, J. Biol. Chem. 250 (1975) 4007.
[2] M. Quadroni, P. James, Electrophoresis 20 (1999) 664.
[3] R. Westermeier, Electrophoresis in Practice, third ed., Wiley-VCH, 2001.
[4] T. Rabilloud, Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods, Springer-Verlag, Berlin, Heidelberg, 2000.
[5] M.P. Molley, E.E. Brzezinski, J. Hang, M.T. McDowell, R.A. Van Bogelen, Proteomics 3 (2003) 1912.
[6] A.W. Dowsey, M.J. Dunn, G.-Z. Yang, Proteomics 3 (2003) 1567.
[7] R.D. Appel, P.M. Palagi, D. Walther, J.R. Vargas, J.-C. Sanchez, F. Ravier, C. Pasquali, D.F. Hochstrasser, Electrophoresis 18 (1997) 2724.
[8] J. Bossinger, M.J. Miller, K.-P. Vo, E.P. Geiduschek, N.-H. Xuong, J. Biol. Chem. 254 (1979) 7986.

[9] K.-P. Vo, M.J. Miller, E.P. Geiduschek, C. Nielsen, A. Olson, N.-H. Xuong, Anal. Biochem. 112 (1981) 258.
[10] P.F. Lemkin, L.E. Lipkin, Comput. Biomed. Res. 14 (1981) 272.
[11] P.F. Lemkin, L.E. Lipkin, Comput. Biomed. Res. 14 (1981) 355.
[12] N.L. Anderson, J. Taylor, A.E. Scandora, B.P. Coulter, N.G. Anderson, Clin. Chem. 27 (1981) 1807.
[13] M.M. Skolnick, S.R. Sternberg, J.V. Neel, Clin. Chem. 28 (1982) 969.
[14] A.D. Olson, M.J. Miller, Anal. Biochem. 169 (1988) 49.
[15] R.D. Appel, D.F. Hochstrasser, M. Funk, J.R. Vargas, C. Pellegrini, A.F. Muller, J.-R. Scherrer, Electrophoresis 12 (1991) 722.
[16] J.I. Garrels, J. Biol. Chem. 264 (1989) 5269.
[17] K.-P. Pleißner, F. Hoffmann, K. Kriegel, C. Wenk, S. Wegner, A. Sahlström, H. Oswald, H. Alt, E. Fleck, Electrophoresis 20 (1999) 755.
[18] Z. Smilansky, Electrophoresis 22 (2001) 1616.
[19] S. Veeser, M.J. Dunn, G.-Z. Yang, Proteomics 1 (2001) 856.
[20] B. Raman, A. Cheung, M.R. Marten, Electrophoresis 23 (2002) 2194.
[21] J. Pánek, J. Vohradský, Electrophoresis 20 (1999) 3483.
[22] T. Voss, P. Haberl, Electrophoresis 21 (2000) 3345.
[23] J.E. Myrick, P.F. Lemkin, M.K. Robinson, K.M. Upton, Appl. Theor. Electrophoresis 3 (1993) 335.
[24] P. Mahon, P. Dupree, Electrophoresis 22 (2001) 2075.
[25] J.C. Nishihara, K.M. Champion, Electrophoresis 23 (2002) 2203.
[26] A.T. Rosengren, J.M. Salmi, T. Aittokallio, J. Westerholm, R. Lahesmaa, T.A. Nyman, O.S. Nevalainen, Proteomics 3 (2003) 1936.
[27] R.D. Appel, J.R. Vargas, P.M. Palagi, D. Walther, D.F. Hochstrasser, Electrophoresis 18 (1997) 2735.
[28] M.C. Pietrogrande, N. Marchetti, F. Dondi, P.G. Righetti, Electrophoresis 23 (2002) 283.
[29] K. Kriegel, I. Seefeldt, F. Hoffmann, C. Schultz, C. Wenk, V. Regitz-Zagrosek, H. Oswald, E. Fleck, Electrophoresis 21 (2000) 2637.
[30] L. Vincent, P. Soille, IEEE Trans. Pattern Anal. Mach. Intell. 13 (1991) 583.
[31] J.X. Yan, J.C. Sanches, L. Tonella, K.L. Williams, D.F. Hochstrasser, Electrophoresis 20 (1999) 738.
[32] M.J. Dutt, K.H. Lee, Electrophoresis 22 (2001) 1627.
[33] L.G. Brown, ACM Comput. Surveys 24 (1992) 325.
[34] J. Salmi, T. Aittokallio, J. Westerholm, M. Griese, A. Rosengren, T.A. Nyman, R. Lahesmaa, O. Nevalainen, Proteomics 2 (2002) 1504.
[35] F.L. Bookstein, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 567.
[36] G. Horgan, A. Creasey, B. Fenton, Electrophoresis 13 (1992) 871.
[37] J.S. Gustafsson, A. Blomberg, M. Rudemo, Electrophoresis 23 (2002) 1731.
[38] C.A. Glasbey, K.V. Mardia, J. R. Statist. Soc. B 63 (2001) 465.
[39] M.M. Skolnick, Clin. Chem. 28 (1982) 979.
[40] P. Vincens, P. Tarroux, Electrophoresis 8 (1987) 100.
[41] F. Hoffmann, K. Kriegel, C. Wenk, Discrete Appl. Math. 93 (1999) 75.
[42] J. Vohradský, Electrophoresis 18 (1997) 2749.
[43] E. Marengo, E. Robotti, V. Gianotti, P.G. Righetti, D. Cecconi, E. Domenici, Electrophoresis 24 (2003) 225.
[44] J. Schultz, D.M. Gottlieb, M. Petersen, L. Nesic, S. Jacobsen, I. Søndergaard, Electrophoresis 25 (2004) 502.
[45] S. Gold, A. Rangarajan, C.-P. Lu, S. Pappu, E. Mjolsness, Pattern Recognit. 31 (1998) 1019.
[46] M. Bennamon, Int. J. Pattern Recognit. Artif. Intell. 17 (2003) 1057 (Editorial).
[47] V. Mäkinen, Parameterized approximate string matching and local-similarity-based point-pattern matching, Ph.D. Thesis, Department of Computer Science, University of Helsinki, Report A-2003-6, 2003.
[48] M. Rogers, J. Graham, R.P. Tonge, Proteomics 3 (2003) 879.
[49] M. Rogers, J. Graham, R.P. Tonge, Proteomics 3 (2003) 887.
[50] W.F. Patton, J. Chromatogr. B 771 (2002) 3.
[51] A. Alban, S.O. Davis, L. Bjorkesten, C. Andersson, E. Sloge, S. Lewis, I. Currie, Proteomics 3 (2003) 36.
[52] K.-P. Pleißner, T. Eifert, S. Buettner, F. Schmidt, M. Boehme, T.F. Meyer, S.H.E. Kaufmann, P.-R. Jungblut, Proteomics 4 (2004) 1305.
[53] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.
[54] S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 9390.
[55] D.K. Han, J. Eng, H. Zhou, R. Aebersold, Nat. Biotechnol. 19 (2001) 946.