

Unsupervised 2D gel electrophoresis image segmentation based on active contours

Michalis A. Savelonas^{*}, Eleftheria A. Mylona and Dimitris Maroulis

Department of Informatics and Telecommunications,
University of Athens, 15784, Panepistimioupolis, Athens, Greece

Abstract

This work introduces a novel active contour-based scheme for unsupervised segmentation of protein spots in two-dimensional gel electrophoresis (2D-GE) images. The proposed segmentation scheme is the first to exploit the attractive properties of the active contour formulation in order to cope with crucial issues in 2D-GE image analysis, including the presence of noise, streaks, multiplets and faint spots. In addition, it is unsupervised, providing an alternate to the laborious, error-prone process of manual editing, which is required in state-of-the-art 2D-GE image analysis software packages. It is based on the formation of a spot-targeted level-set surface, as well as of morphologically-derived active contour energy terms, used to guide active contour initialization and evolution, respectively. The experimental results on real and synthetic 2D-GE images demonstrate that the proposed scheme results in more plausible spot boundaries and outperforms all commercial software packages in terms of segmentation quality.

Keywords: Segmentation; Active contours; 2D-gel electrophoresis images.

^{*} Corresponding author

Tel.: +30-210-7275317

Fax: +30-210-7275333

E-mail addresses: {m.savelonas, e.mylona, d.maroulis}@di.uoa.gr

1 Introduction

Protein expression is highly indicative of various pathological conditions ranging from neoplasms and tumors to infectious diseases and genetic disorders. In this light, protein patterns of normal and diseased origin are compared in order to allow the identification of possible differences in protein expression. The platform utilized for protein mapping is called two-dimensional gel electrophoresis (2D-GE) [1].

In 2D-GE, an indicative portion of the total protein component of a cell is resolved and information about different post translational modifications attributed to proteins is provided. Proteins travel across the gel in two dimensions: horizontal and vertical, which reflect protein isoelectric point and protein molecular weight, respectively [2],[3]. They are separated according to the isoelectric point by applying a pH gradient to the gel and an electric potential across the gel, which causes each charged protein to migrate towards the oppositely charged electrode. The accumulated amounts of separated proteins are detected either by radioactive labeling or staining techniques. The results of gel electrophoresis are captured in digital images, where proteins are represented as spots over a grey level surface. The amount of each migrated protein can be estimated by the cumulative intensity of the associated spot region. The computational analysis of protein content on 2D-GE images is a challenging pattern recognition task, which involves several layers of processes including 2D-GE image segmentation and quantification.

2D-GE image segmentation is the process of separating protein spots from 2D-GE image background. Various issues arise in this process including the presence of noise as well as of dust particles, fingerprints and cracks on the gel surface. In addition, illumination may result in inhomogeneous background intensity, whereas protein expression ranges from faint to saturated spots. Moreover, protein mixtures from cells, tissues or biological fluids comprise more than 10,000 proteins. The mixture complexity obstructs proteins migration, leading to complex regions containing overlapping spots. Such “multiplets” tend to occupy a large portion of the gel surface impeding 2D-GE segmentation.

Several methods have been suggested to tackle with 2D-GE image analysis such as stepwise thresholding

[4], edge detection [5] and watersheds [6]. Stepwise thresholding applies an increasing threshold on the 2D-GE image of interest, starting from the lowest intensity level which can be associated with a protein spot. As the applied threshold increases, each connected image area may be split into multiple connected sub-areas. This process is iterative and stops when no more splits are possible. The segmentation result is determined by the connected sub-areas remaining in the last split. Edge detection methods aim to identify discontinuities in image intensity, often associated with protein spot boundaries. Both stepwise thresholding and edge detection methods are highly sensitive to noise, artifacts, non-uniform background and overlapping spot clustering [7], whereas manual editing may be required in the case of edge detection [8]. Watershed methods model a 2D-GE image as a landscape where rain falls downhill formulating pools around each local intensity extremum. Areas collecting the water in each pool are called catchment basins and can be associated with protein spots. Although this approach copes with the presence of noise, artifacts and non-uniform background [9], it calls for additional post-processing since all pixels in the image are assigned to a catchment basin, resulting in over-segmentation [10]. Aiming to cope with this issue, Kim et al. [11] introduced a hybrid 2D-GE image segmentation approach based on watersheds and stepwise thresholding. However, the background removal process incorporated in this hybrid approach cannot cope with the presence of faint spots. Similar variations have been introduced for the segmentation of cell images [12]-[14]. 2D-GE image analysis software packages, such as PDQuest (Bio-Rad) [15], Melanie (GeneBio/GE Healthcare) [16] and Delta2D (Decodon) [17], are dominant in the research field of gel analysis. However, such software packages are highly parametric and demonstrate a notable output variance [18]. Furthermore, the results obtained by each software package are manually edited by gel analysts so as to eliminate false positive spots, reconsider false negative spots or correct the elliptical or circular boundaries used to define the spot area.

Active contours [19] have been the dominant segmentation approach in the last two decades, as they are self-adapting and lead to continuous curves, without requiring edge-linking operations. Moreover, the

inherent continuity and smoothness of active contours cope with the presence of noise, gaps, and other irregularities in object boundaries. Furthermore, when formulated using level-sets [20],[21], active contours are able to adapt to topological changes such as contour splitting or merging [22]-[24]. This latter attribute is of particular importance in cases of 2D-GE images containing a few hundred up to several thousands of protein spots. A first attempt addressing the application of active contours on 2D-GE images appeared in 2008 [25], whereas our preliminary works introducing some of the ideas incorporated in the proposed segmentation scheme can be found in [26]-[29]. However, [25] involved a straightforward application of the Chan-Vese model which cannot cope with the presence of streaks, multiplets and faint spots. On the other hand, [26]-[28] aimed at protein spot detection and not at 2D-GE image segmentation, whereas [29] introduced an initial version of the idea presented in this work with respect to boundary identification of overlapping spots without addressing the presence of streaks and faint spots.

In this work, a novel active contour-based scheme is proposed for the segmentation of protein spots in 2D-GE images. To the best of our knowledge, this is the first complete segmentation scheme exploiting the attractive properties of the active contour formulation in order to cope with crucial issues in 2D-GE image analysis, including the presence of noise, streaks, multiplets and faint spots. In addition, it is unsupervised, providing an alternate to the laborious, error-prone process of manual editing, which is required in state-of-the-art 2D-GE image analysis software packages. It comprises of four main processes namely: (a) a detection process capable of identifying boundaries of spot overlap in regions occupied by multiplets, based on the observation that such boundaries are associated with local intensity minima, (b) histogram adaptation and morphological reconstruction so as to avoid unwanted amplifications of noise and streaks, as well as to facilitate the identification of faint spots, (c) a contour initialization process aiming to form a level-set surface initializing the subsequent level-set evolution, based on the observation that protein spots are associated with regional intensity maxima, and (d) a level-set evolution process guided by region-based energy terms determined by image intensity as well as by information derived from the previous processes.

The remainder of this paper is organized in four sections. Section 2 provides the theoretical background of the Chan-Vese active contour and mathematical morphology, respectively whereas Section 3 presents the main components of the proposed scheme. Section 4 demonstrates the experimental results on real and synthetic 2D-GE images, as well as comparisons with PDQuest 8.0.1, Melanie 7 and Delta2D software packages. Finally, conclusions and future perspectives of this work are discussed in Section 5.

2 Background

2.1 The Chan-Vese active contour on 2D-GE images

The Chan-Vese active contour [30] is a region-based level-set model which is particularly suited to 2D-GE image segmentation due to its robustness to the presence of noise, its topological adaptability, as well as its capability of detecting smooth boundaries or boundaries that are not defined by gradient, as is the case with protein spots. The mathematical formulation of the Chan-Vese active contour adopts the reduced case of the Mumford-Shah problem [31], resulting in the following evolution equation:

$$\frac{\partial \varphi}{\partial t} = \delta(\varphi) [\mu \cdot \operatorname{div} \left(\frac{\nabla \varphi}{|\nabla \varphi|} \right) - \lambda_1^+ (u_1 - c_1^+)^2 + \lambda_1^- (u_1 - c_1^-)^2] \quad (1)$$

where u_1 is a two-dimensional image, $\varphi(x, y)$ is the level-set function, c_1^+, c_1^- are the respective average intensities, δ is Dirac delta function, t is the artificial time parameterizing the descent direction and $\mu, \lambda_1^+, \lambda_1^- > 0$ are weighting parameters. The average intensities c_1^+, c_1^- are iteratively updated as:

$$c_1^+(\varphi) = \frac{\int_{\Omega} u_1(x, y) H(\varphi(x, y)) dx dy}{\int_{\Omega} H(\varphi(x, y)) dx dy} \quad c_1^-(\varphi) = \frac{\int_{\Omega} u_1(x, y) (1 - H(\varphi(x, y))) dx dy}{\int_{\Omega} (1 - H(\varphi(x, y))) dx dy} \quad (2)$$

where H is the Heaviside function.

Despite the aforementioned advantages, the Chan-Vese active contour still fails to accurately segment 2D-GE images containing considerable amounts of overlapping or faint protein spots. Figure 1 illustrates an

example of two overlapping protein spots and the segmentation results obtained by a straightforward application of the Chan-Vese active contour. It is evident that the Chan-Vese active contour merges overlapping spot boundaries. Moreover, the convergence of the Chan-Vese active contour is not completely insensitive to initialization [32].

[Figure 1]

2.2 *Mathematical morphology*

Mathematical morphology [33] is a well-known image analysis approach, which can be applied for the extraction or suppression of image components of interest by designing a suitable structuring element (SE). Morphological operations, including dilation and erosion, are capable of preserving topological properties such as connectivity and homotopy [34], whereas they are suitable for detecting intensity peaks associated with protein spots in 2D-GE images.

The shape and size of the SE is often selected in accordance with the shape of the objects of interest [35]. In the case of 2D-GE images, where the dominant shape of protein spots is circular, a disk-shaped SE may lead to better results. As it comes to size, a large disk tends to ignore most of regional intensity maxima. Based on these considerations, the SE is selected disk-shaped and the radius r is selected small in order to minimize missed regional intensity maxima.

A regional intensity maximum M at elevation t is defined by:

$$\begin{cases} I(p) = t, \forall p \in M \\ I(p) < t, \forall p \in \delta_{SE}(M) \setminus M \end{cases} \quad (3)$$

where p is the pixel location, $I(p)$ is the intensity of p and δ_{SE} is the region generated by the dilation of M according to the SE.

3 Proposed 2D-GE image segmentation scheme

The proposed segmentation scheme comprises of four main processes for: (a) detection of multiplets, (b) histogram adaptation and morphological reconstruction, (c) contour initialization and (d) level-set evolution.

3.1 Detection of local intensity minima in multiplets

It can be observed that the region of overlap between two protein spots is associated with local intensity minima, with respect to a particular direction. Figure 2 illustrates this point by three-dimensional representations of protein spot intensities, in cases of partly overlapped (Fig. 2a) and highly overlapped (Fig. 2b) protein spots. This observation motivated us to incorporate information derived by such local intensity minima in the proposed 2D-GE image segmentation scheme.

[Figure 2]

The original image is pre-processed with a $k \times k$ median filter [36] aiming to reduce the side effects of noise on the following processes. The pre-processed image is scanned with parallel straight-line segments of variable lengths and multiple directions (Fig. 3) so as to facilitate the detection of local intensity minima, associated with each particular direction.

[Figure 3]

Local intensity minima are identified for each parallel straight-line segment, however the ones which are eventually selected conform to the following two criteria: a) intensity value exceeds a threshold value T_1 and b) intensity value is a global minimum over a square sub-segment of width exceeding a minimum value w . These criteria are imposed to exclude local intensity minima associated with background clutter. Figure 4 illustrates: a) a real 2D-GE image, b) the detection results obtained by the local intensity minima process, with each minimum marked as black and a_1 - b_1 respective sub-images as marked in a and b. In Fig. 4b₁ it is evident that the detection process actually identifies boundaries of spot overlap. Therefore, alterations in the

pre-processing techniques as well as further manual editing are not required.

[Figure 4]

3.2 *Image enhancement and morphological reconstruction*

A popular histogram equalization variant called contrast-limited adaptive histogram equalization (CLAHE) [37] is utilized to enhance the segmentation performance of the proposed scheme with respect to the presence of faint spots in 2D-GE images. CLAHE involves a grayscale transformation function which has been effectively applied on various medical imaging modalities including mammographic [38] and chest CT [39] imaging. The core idea is to adaptively enhance image contrast in a local fashion, contrary to the original histogram equalization which is uniformly applied on the entire image.

CLAHE separately applies histogram equalization on $q \times q$ small non-overlapping image regions called tiles. The histogram of each transformed tile approximates the uniform distribution, which results in amplification of faint regions such as the faint protein spots in 2D-GE images. The local nature of CLAHE prevents unwanted amplifications of noise and streaks, as opposed to the original histogram equalization which has the same amplifying effect on noise and artifacts as on faint spots. In addition, CLAHE imposes a constraint on the resulting contrast providing a mechanism to cope with a possible over-saturation of the resulting image. This constraint can be adjusted by a parameter h , called clip limit. The clip limit h determines the maximum number of pixels which are allowed to occupy a bin in the resulting histogram. In cases of over-saturation, where certain histogram bins are occupied by more than $h \times (2^{\text{gray level depth}} - 1)$ pixels, the excessive amount of pixels is redistributed over the rest of the histogram. The neighboring transformed tiles are then merged using bilinear interpolation to reduce artificially induced boundaries and the pixel intensity values are updated in accordance with the adapted histograms [40].

Figure 5 illustrates the images resulted from the application of: a) histogram equalization and b) CLAHE, on the original 2D-GE image of Fig. 4b. A sub-image of this original 2D-GE image is illustrated in c,

whereas the corresponding sub-images of Fig. 5a and b are magnified in Fig. 5a₁ and 5b₁, respectively. It can be observed that both techniques amplify spots which were faint in the original 2D-GE image; however CLAHE avoids unwanted amplifications of noise and streaks, which is not the case with the plain histogram equalization. Figure 6 illustrates the histograms of: a) the original 2D-GE image illustrated in Fig. 4b, as well as the histograms of the images illustrated in b) Fig. 5a and c) Fig. 5b. It can be observed that the histogram of the image resulted by the application of CLAHE is much denser than the one generated by plain histogram equalization, indicating that the former maintains much more detailed image-related information.

It should be noted that the application of CLAHE could not benefit the detection of local intensity minima described in Section 3.1, as well as the contour initialization process which is described in Section 3.3. Accordingly, both processes are applied on the original 2D-GE image. This can be justified by considering that the clipping involved in CLAHE redistributes pixels over the histogram, introducing intensity minima which are not necessarily associated with spot boundaries, as well intensity maxima which are not necessarily associated with spots.

[Figure 5]

[Figure 6]

The protein spot regions depicted on the enhanced image generated from CLAHE technique are still characterized by intensity inhomogeneity which would affect the subsequent active contour evolution. Another morphological processing step is performed in order to cope with this issue. The enhanced image is binarized according to a threshold value T_2 . However, the protein spot regions of the binary image contain holes as a result of intensity inhomogeneity. The flood-fill morphological operation [33] is applied so as to eliminate such holes. This morphological operation alters the connected background pixels to foreground pixels until it reaches the object boundaries.

Figure 7 illustrates the results obtained by the flood-fill morphological operation on: a) the original 2D-GE image illustrated in Fig. 4b, and b) on the enhanced image of Fig. 5b, which is generated by the application of CLAHE. A sub-image of the original 2D-GE image is illustrated in c), whereas a_1 and b_1 are the corresponding sub-images of Fig. 7a and Fig. 7b, respectively. Missing regions in Fig. 7a₁, as compared to Fig. 7b₁, correspond to faint protein spots. It is evident that the utilization of CLAHE is essential, since most faint spots are missed when CLAHE is omitted. The obtained binarized image represents protein spots, including faint ones, as well as the boundaries of spot overlap in regions occupied by multiplets. This indispensable information is incorporated in contour evolution, as described in Section 3.4.

[Figure 7]

3.3 *Contour initialization*

The Chan-Vese active contour is not completely insensitive to initialization [32]; therefore it is essential to initialize the level-set function so that the associated zero levels approximate the actual protein spots. Emerging from the observation that regional intensity maxima of a 2D-GE image are associated with protein spots, the proposed initialization process aims to detect such maxima in order to construct a level-set surface of multiple cones centered at maxima positions. This surface can serve as a spot-targeted initialization of the level-set function. Such an initialization process is particularly important within the context of the proposed scheme, in a sense that it provides the capability of unsupervised segmentation. In this light, this process is one of the novel elements of the proposed scheme, since it extends a straightforward active contour application, which would have required supervised initialization so as to avoid sub-optimal segmentation results. It should be noted that the level-set function is initialized on the 2D-GE image instead of the binarized image described in Section 3.2, since regional intensity maxima are not maintained in the latter image.

Aiming to include salient intensity maxima positions associated with protein spots and avoid spurious ones associated with background noise peaks, we impose the following constraints on the selection of regional intensity maxima:

- a) intensity should be equal to the maximum intensity value over an $m \times m$ adjacent region.
- b) every pixel over a $z \times z$ square neighborhood of each selected regional intensity maximum should have intensity which exceeds a threshold value T_3

The positions of selected regional intensity maxima are used as centers of cones forming the surface of the initial level-set function. Apart from cone centers, the proposed initialization process determines the zero-level regions associated with each cone. A disk-shaped SE (see Section 2.2) is used to form these regions, considering that the dominant shape of protein spots in 2D-GE images is approximately cyclical.

The original 2D-GE image is dilated with a disk-shaped SE. SE radius r is set according to the results of preliminary experimentation on 2D-GE images, which indicate that a certain radius value minimizes the detection of false negative protein spots whereas it allows the detection of local intensity maxima associated with small spots even in cases where they overlap with larger spots in complex regions. This radius value occurs to be smaller than the typical size of a protein spot, which ranges from 20 to 100 pixels.

Figure 8 illustrates a three-dimensional representation of the level-set surface of multiple cones obtained by the application of the proposed initialization process on a real 2D-GE image.

[Figure 8]

3.4 Contour evolution

Aiming to enhance segmentation performance, contour evolution is initialized by the spot-targeted level-set surface generated by the previous initialization process. In addition, the active contour evolves in separate $g \times g$ image sub-regions, which are centered at the cone centers of the level-set surface. The active contour converges according to the following equation:

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[\mu \cdot \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \lambda_1^+ (u_1 - c_1^+)^2 + \lambda_1^- (u_1 - c_1^-)^2 - \lambda_2^+ (u_2 - c_2^+) + \lambda_2^- (u_2 - c_2^-) \right] \quad (4)$$

where u_1, u_2 are a 2D-GE image, and the binarized image which is the output of morphological processing described in Section 3.2, respectively. In addition, c_1^+, c_2^+ and c_1^-, c_2^- are the average foreground and background intensities of u_1 and u_2 , calculated by Eq. (2), whereas λ_1^+, λ_2^+ and λ_1^-, λ_2^- are the weights for the regularizing and fitting terms of u_1 and u_2 , respectively. Equation (4) describing contour evolution of the proposed scheme extends Eq. (1) in the sense that it encompasses information derived by: 1) the 2D-GE image u_1 , 2) the binarized image u_2 obtained by the application of adaptive histogram equalization and morphological processing of the original 2D-GE image, as described in Section 3.2. The latter information is essential to identify the presence of faint spots as well as the boundaries of spot overlap in regions occupied by multiplets.

4 Results

The experimental evaluation of the proposed scheme has been conducted on a dataset of 16 real digital grayscale 2D-GE images provided by the Biomedical Research Foundation of the Academy of Athens, as well as on a dataset of 30 synthetic 2D-GE images, so as to facilitate qualitative and quantitative comparisons with state-of-the-art 2D-GE image analysis software packages. The size of each real and synthetic 2D-GE image used was approximately 2000×3000 and 1500×2000 pixels respectively, whereas image gray-level depth of both image types was 16-bit. The proposed scheme has been implemented in Matlab R2009b and executed on a 3.2 GHz Intel Pentium workstation.

Parameter tuning was based on preliminary experimentation, which resulted in the values presented in Table I. The preliminary experiments were performed on three pilot 2D-GE images, whereas the search on the parameter space was guided by the following considerations:

- the window size k and the width w of the square sub-segment considered in local intensity minima detection, as well as the size z of the square neighborhood considered in contour initialization process, were all set to 3. This value is the smallest value for these three parameters, whereas higher values of k resulted in missing spots in the contour initialization process and higher values of both w and z resulted in slight reduction of the obtained segmentation quality,
- thresholds T_1, T_2, T_3 were experimentally identified as 150, 160 and 75, since these values approximate: 1) the upper extreme of the intensity range of the background clutter, 2) the lower extreme of the intensity range of faint spots on the images resulted from histogram equalization and 3) the lower extreme of the intensity range of faint spots on the original 2D-GE images, respectively. Perturbations of T_1, T_2, T_3 within the ranges [147,152], [158,162] and [72,77], resulted in insignificant variations of the obtained segmentation quality,
- tile size q considered for CLAHE was experimentally identified as 40, since this values approximates the typical size of a faint protein spot. Perturbations of q within the range [37,42] resulted in insignificant variations of the obtained segmentation quality,
- clip limit h is set to 0.01 since this order of magnitude reduces over-saturation and lead to the optimal segmentation quality in all pilot 2D-GE images. Perturbations of h within the range [0.006,0.03] resulted in insignificant variations of the obtained segmentation quality,
- size m of adjacent regions and radius r considered in contour initialization process as well as image sub-region size g considered in contour evolution process were experimentally identified as 20, 4 and 50, since these values approximate: 1) the lower extreme of protein spot sizes and consider salient intensity maxima associated with protein spots, 2) the lower extreme of protein spot radii and allow the detection of regional intensity maxima in cases of small spots overlapping with larger spots in multiplets and 3) the average size of a typical protein spot. Perturbations of m, r and g within the ranges [17,24], [3,5] and [42,59] resulted in insignificant variations of the obtained segmentation quality,

- following relevant literature [30], the weights of the energy terms λ_1^+ , λ_1^- , λ_2^+ and λ_2^- were set to 1 whereas the weight μ was adjusted to as $0.006 \cdot 255^2$, since this value lead to the optimal segmentation quality in all pilot 2D-GE images. Perturbations of μ within the range $[0.003 \cdot 255^2, 0.009 \cdot 255^2]$ resulted in insignificant variations of the obtained segmentation quality.

The variations in segmentation quality are considered as insignificant when the values of the associated segmentation quality measures (i.e. volumetric overlap and volumetric error, as defined in Eq. 8), as derived for each one of the three pilot 2D-GE images are overlapping. The latter occurs when the values of the segmentation quality measure derived for a pilot 2D-GE image are within the ranges defined by the mean values and the standard deviations of the same measure, as derived for the other two pilot 2D-GE images. It should be pointed out that parameter tuning is performed *once* on a small number of pilot 2D-GE images generated with a certain experimental setup (pH, staining etc) and the resulting parameter values can be used for all 2D-GE images generated with the same setup. On the contrary, state-of-the-art software packages require parameter tuning *for each single* 2D-GE image, as confirmed by expert biologists.

Table 1

Parameter values

Detection of local intensity minima in multiplets	$k = 3$	$T_1 = 150$	$w = 3$				
Image enhancement and morphological reconstruction	$q = 40$	$h = 0.01$	$T_2 = 160$				
Contour initialization	$m = 20$	$z = 3$	$T_3 = 75$	$r = 4$			
Contour evolution	$g = 50$	$\mu = 0.006 \cdot 255^2$	$\lambda_1^+ = 1$	$\lambda_1^- = 1$	$\lambda_2^+ = 1$	$\lambda_2^- = 1$	

Figure 9 illustrates example segmentation results obtained by the application of the proposed scheme, as well as of PDQuest 8.0.1, Melanie 7 and Delta2D image analysis commercial software packages, on a real

2D-GE image. It should be noted that the output images resulting from the application of the software packages varied with respect to size and resolution. The software packages were applied on inverted versions of the 2D-GE images, whereas parameter settings and calibrations involved were performed by expert biologists, following their experience.

[Figure 9]

It is evident that the proposed scheme results in more plausible spot boundaries (Fig. 9a₁) than all three image analysis software packages, namely PDQuest 8.0.1 (Fig. 9b₁), Melanie 7 (Fig. 9c₁) and Delta2D (Fig. 9d₁). PDQuest 8.0.1 results in elliptical boundaries which do not correspond to the irregular shape of the actual spot boundaries, whereas such elliptical boundaries tend to include background regions. In the cases of Melanie 7 and Delta2D, the segmentation results obtained suffer from over-segmentation and are subject to laborious, error-prone and time-consuming correction process by the expert biologists.

In order to quantitatively evaluate the proposed scheme, experiments were performed on the set of synthetic images generated by the synthetic 2D-GE image generation software, developed by the Real-time Systems & Image Analysis Lab. Figure 10 illustrates an example of a synthetic 2D-GE image, as well as the corresponding ground truth. Such a synthetic image is populated by approximately 200 spots, following beta distribution. As a result of trial-and-error experimentation, parameters a and b of the beta function were set to 4 and 3 respectively, resulting in spatial frequency of singlet and multiplet occurrence which emulates real 2D-GE images. Synthetic background emulates inhomogeneity, streaks and clutter, which characterize the background of real 2D-GE images.

[Figure 10]

The intensity profile of each spot is chosen flat top in order to emulate the saturation characterizing actual protein spots and is defined by:

$$I(x, y) = \begin{cases} 1, & r \leq r_0 \\ \cos^2\left(\frac{r-r_0}{2\pi\sigma_\phi}\right), & r_0 < r \leq r_0 + \sigma_\phi \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

where r_0 is the radius of the flat top, r is the Euclidean distance from the center of the spot and σ_ϕ^2 is an angle-dependent variance coefficient:

$$\sigma_\phi^2 = \frac{(r_0 + \sigma_x)(r_0 + \sigma_y)r}{\sqrt{(r_0 + \sigma_x)^2(x - x_0)^2 + (r_0 + \sigma_y)^2(y - y_0)^2}} - r_0 \quad (6)$$

where σ_x and σ_y are the variance coefficients along the primary axes. Figure 11 illustrates example segmentation results obtained by the proposed scheme, as well as by PDQuest 8.0.1, Melanie 7 and Delta2D software packages on a synthetic 2D-GE image.

[Figure 11]

The segmentation results are quantified according to the spot volume V , as defined in [16]:

$$V = \sum_{x,y \in \text{region}} I(x, y) \quad (7)$$

where $I(x,y)$ is the intensity value of pixel (x,y) .

Comparison of the segmentation results with the corresponding ground truth image, as generated by the 2D-GE image simulation software allows the categorization of each pixel in one of the following four region types: “actual spot region (ASR)”, “false spot region (FSR)”, “false background region (FBR)” and “actual background region (ABR)”.

The spot volumes which are calculated according to Eq. (7) for the above four cases of regions, correspond to the “actual spot volume” (ASV), “false spot volume” (FSV), “false background volume” (FBV) and “actual background volume” (ABV), respectively. The segmentation performances are

quantitatively evaluated in terms of *volumetric overlap* vo and *volumetric error* ve , which are defined as follows:

$$vo = \frac{ASV}{ASV + FBV}, ve = \frac{FSV}{ASV + FBV} \quad (8)$$

Table II presents the results obtained by the proposed scheme, as well as by PDQuest 8.0.1, Melanie 7 and Delta2D software packages. Figure 12 provides a visualization of the results of Table II. It is evident that the proposed scheme outperforms all three software packages in terms of vo and ve . In particular, the E obtained by the proposed scheme is approximately 3-4 times smaller than the one obtained by the software packages, indicating that it is much more effective in avoiding the identification of FSR. Moreover, the proposed scheme demonstrates a remarkably lower variance in both performance measures, as a result of its robustness over streaks, multiplets and faint spots.

Table 2

Segmentation results

	Proposed Scheme	PDQuest 8.0.1	Melanie 7	Delta2D
vo	92.0±1.2%	80.2±4.6%	86.5±3.2%	82.4±3.6%
ve	20.0±3.2%	83.1±8.9%	55.0±6.7%	64.3±7.6%

[Figure 12]

5 Conclusions

In this work, a novel active contour-based scheme is proposed for unsupervised segmentation of 2D-GE images. The proposed segmentation scheme is the first to exploit the attractive properties of the active contour formulation in order to cope with crucial issues in 2D-GE image analysis, including the presence of noise, streaks, multiplets and faint spots. It incorporates: (a) a detection process capable of identifying boundaries of spot overlap in regions occupied by multiplets, based on the observation that such boundaries

are associated with local intensity minima, (b) histogram adaptation and morphological reconstruction so as to avoid unwanted amplifications of noise, streaks and facilitate the identification of faint spots, (c) a contour initialization process aiming to form a level-set surface initializing the subsequent contour evolution, based on the observation that protein spots are associated with regional intensity maxima, and (d) a contour evolution process guided by region-based energy terms determined by image intensity as well as by information derived from the previous processes of the proposed scheme.

The experimental evaluation of the proposed scheme has been conducted on datasets of both real and synthetic 2D-GE images, so as to facilitate quantitative comparisons with state-of-the-art 2D-GE image analysis software packages, including PDQuest 8.0.1, Melanie 7 and Delta2D. As it can be derived by the experimental results, the proposed scheme: (a) is capable of identifying spot boundaries within regions occupied by multiplets, (b) is capable of identifying boundaries of faint spots, (c) copes with the presence of noise, as a result of the region-based formulation of the energy terms in contour evolution equation, (d) results in more plausible spot boundaries than PDQuest 8.0.1, Melanie 7 and Delta2D 2D-GE image analysis software packages as it can be observed on the segmentation results on both real and synthetic 2D-GE images, (e) outperforms all three 2D-GE image analysis software packages in terms of segmentation quality measures, calculated from the segmentation results obtained on synthetic 2D-GE images, and (f) is unsupervised, providing an alternate to the laborious, error-prone and time-consuming process of manual editing, which is required in state-of-the-art 2D-GE image analysis software packages.

Future perspectives of this work involve integration of the proposed scheme within a 2D-GE image analysis system, applicable in everyday practice of biologists.

Acknowledgement

This work has been co-financed by the European Union (European Social Fund-ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic

Reference Framework (NSRF)- Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund. We would like to thank the Biomedical Research Foundation of the Academy of Athens for the provision of real 2D-GE images as well as segmentation results obtained by Melanie 7 software package. We would also like to thank expert biologists M. Makridakis and M. Aivaliotis for the provision of segmentation results obtained by PDQuest 8.0.1 and Delta2D software packages respectively on real 2D-GE images, as well as Dr. S. Kossida and Dr. A. Vlahou for their constructive comments on the obtained results. Finally, we would particularly like to thank the reviewers for their fruitful comments and suggestions.

References

- [1] A.W. Dowsey, M.J. Dunn, G.Z. Yang, The role of bioinformatics in two-dimensional gel electrophoresis, *Proteomics* 3 (8) (2003) 1567-1596.
- [2] K. Rohr, P. Cathier, S. Wölz, Elastic registration of electrophoresis images using intensity information and point landmarks, *Pattern Recognition* 37 (2004) 1035-1048.
- [3] M. Berth, F.M. Moser, M. Kolbe, J. Bernhardt, The state of the art in the analysis of two-dimensional gel electrophoresis images, *Appl. Microb. Biotechnol.* 76 (6) (2007) 1223-1243.
- [4] J.J. Tyson, R.H. Haralick, Computer analysis of two-dimensional gels by a general image processing system, *Electrophoresis* 7 (1986) 107-113.
- [5] P.F. Lemkin, L.E. Lipkin, 2-D electrophoresis gel data-base analysis - aspects of data structures and search strategies in gellab, *Electrophoresis* 4 (1) (1983) 71-81.
- [6] K.P. Pleissner, F. Hoffman, K. Kriegel, C. Wenk, S. Wegner, A. Sahlstrom, H. Oswald, H. Alt, E. Fleck, New algorithmic approaches to protein spot detection and pattern matching in two-dimensional electrophoresis databases, *Electrophoresis* 20 (4-5) (1999) 755-765.

- [7] P. Cutler, G. Heald, I.R. White, J. Ruan, A novel approach to spot detection for two-dimensional gel electrophoresis images using pixel value collection, *Proteomics* 3 (4) (2003) 392-401.
- [8] K. Takahashi, Y. Watanabe, M. Nakazawa, A. Konagaya, Fully automated spot recognition and matching algorithms for 2-D gel electrophoretogram of genomic DNA, *Genome Inf. Ser. Workshop 9* (1998) 161-172.
- [9] M. B. Rye, Image segmentation and multivariate analysis in two-dimensional gel electrophoresis, PhD Thesis, Norwegian University of Science and Technology, Faculty of Natural Sciences and Technology, Department of Chemistry, Trondheim, Norway, 2007.
- [10] L. Vincent, P. Soille, Watersheds in digital spaces: an efficient algorithm based on immersion simulations, *IEEE Trans. Patt. Anal. and Mach. Intel.*, 13 (6) (1991) 583-598.
- [11] Y. Kim, J. Kim, Y. Won, Y. In, Segmentation of protein spots in 2-D gel electrophoresis images with watershed using hierarchical threshold, *Lecture Notes in Computer Science* 2869 (2003) 389-396.
- [12] V. Barra, Robust segmentation and analysis of DNA microarray spots using an adaptive split and merge algorithm, *Comput. Methods Programs Biomed.* 81 (2006) 174-180.
- [13] M.A. Zapala, D.J. Lockhart, D.G. Pankratz, A.J. Garcia, C. Barlow, Software and methods for oligonucleotide and cDNA array data analysis, *Genome Biol.* 3 (6) (2002) software 0001.1-software 0001.9.
- [14] D. Verellen, W. De Neve, F. Van den Heuvel, M. Coghe, O. Louis, G. Storme, On-line portal imaging: Image quality defining parameters for pelvic fields--a clinical evaluation, *Int. J. Radiat. Oncol. Biol. Phys.* 27 (1993) 945-952.
- [15] J.I. Garrels, The Quest system for quantitative analysis of two-dimensional gels, *Journal of Biological Chemistry* 264 (9) (1989) 5269-5282.

- [16] R.D. Appel, J.R. Vargas, P.M. Palagi, D. Walther, D.F. Hochstrasser, Melanie II – A third generation software package for analysis of two-dimensional electrophoresis images: II. algorithms, *Electrophoresis* 8 (15) (1997) 2735-2748.
- [17] <http://www.decodon.com>
- [18] B.N. Clark, H.B. Gutstein, The myth of automated, high-throughput two-dimensional gel analysis, *Proteomics* 8 (6) (2008) 1197-1203.
- [19] M. Kass, A. Witkin, D. Terzopoulos, Snakes - Active contour models, *Int. J. Comp. Vis.* 1 (4) (1987) 321-331.
- [20] S. Osher, J.A. Sethian, Fronts propagating with curvature-dependent speed – algorithms based on Hamilton-Jacobi formulations, *J. Comp. Phys.* 79 (1) (1988) 12-49.
- [21] Y.-T. Chen, A level-set method based on the Bayesian risk for medical image segmentation, *Pattern Recognition* 43 (2010) 3699-3711.
- [22] Z. Ying, L. Guangyao, S. Xiehua, Z. Xinmin, Geometric active contours without re-initialization for image segmentation, *Pattern Recognition* 42 (2009) 1970-1976,
- [23] W. Fang, K.L. Chang, Incorporating shape prior into geodesic active contours for detecting partially occluded object, *Pattern Recognition* 40 (2007) 2163-2172.
- [24] P. Horvath, I. Jermyn, Z. Kato, J. Zerubia, A higher-order active contour model of a “gas of circles” and its application to tree crown extraction, *Pattern Recognition* 42 (5) (2009) 699--709.
- [25] P. Tsakanikas, E.S. Manolakos, Active contours based segmentation of 2DGE proteomics images, *Proc. European Signal Processing Conference (EUSIPCO)*, 2008.
- [26] M. Savelonas, E. Mylona, D. Maroulis, A level set approach for proteomics image analysis, *Proc. European Signal Processing Conference (EUSIPCO)*, 2010, pp. 1229-1233.

- [27] E.A. Mylona, M.A. Savelonas, D. Maroulis, A. Vlahou, M. Makridakis, Protein spot detection in 2D-GE images using morphological operators, Proc. IEEE International Symposium on Computer-Based Medical Systems (CBMS), 2010.
- [28] E. Mylona, M. Savelonas, D. Maroulis, A two-stage active contour-based scheme for spot detection in proteomics images, Proc. IEEE International Conference on Information Technology Applications in Biomedicine (ITAB), 2010.
- [29] M. Savelonas, E. Mylona, D. Maroulis, Segmentation of two-dimensional gel electrophoresis images containing overlapping spots, Proc. IEEE International Conference on Information Technology Applications in Biomedicine (ITAB), 2009.
- [30] T.F. Chan, L.A. Vese, Active contour without edges, IEEE Trans. Im. Proc. 10 (2) (2001) 226-277.
- [31] D. Mumford, J. Shah, Optimal approximation by piecewise smooth functions and associated variational problems, Comm. Pure Appl. Math. 42 (1989) 577-685.
- [32] S.H. Lee, J.K. Seo, Level set-based bimodal segmentation with stationary global minimum, IEEE Trans. Im. Proc. 15 (9) (2006) 2843-2852.
- [33] P. Soille, Morphological image analysis-principles and applications, Springer, Berlin, 1999.
- [34] P. Dokládál, I. Bloch, M. Couprie, D. Ruijters, R. Urtasun, L. Garnero, Topologically controlled segmentation of 3D magnetic resonance images of the head by using morphological operators, Pattern Recognition 36 (2003) 2463-2478.
- [35] E.R. Urbach, M.H.F. Wilkinson, Efficient 2-D grayscale morphological transformations with arbitrary flat structuring elements, IEEE Trans. Im. Proc. 17 (2008) 1-8.
- [36] D.T. Lin, Autonomous sub-image matching for two-dimensional electrophoresis gels using MaxRST algorithm, Im. Vis. Comp., In Press, Corrected Proof, 2010.
- [37] S.M. Pizer, E.P. Amburn, J.D. Austin, Adaptive histogram equalization and its variations, Comp. Vis. Graph. Im. Proc. 39 (1987) 355-368.

- [38] A.P. Stefanoyannis, L. Costaridou, S. Skiadopoulos, G. Panayotakis, A digital equalization technique improving visualization of dense mammary gland and breast periphery in mammography, *Eur. J. Radiol.* 45 (2003) 139-149.
- [39] L.M. Fayad, Y. Jin, A.F. Laine, Y.M. Berkmen, G.D. Pearson, B. Freedman, R.V. Heertum, Chest CT window settings with multiscale adaptive histogram equalization: pilot study, *Radiology* 223 (2002) 845-852.
- [40] E.D. Pisano, S. Zong, B.M. Hemminger, M. DeLuca, R.E. Johnston, K. Muller, M.P. Braeuning, S.M. Pizer, Contrast limited adaptive histogram equalization image processing to improve the detection of simulated speculations in dense mammograms, *J. Digit. Im.* 11 (1998) 193-200.

Figure Captions

Fig. 1. Example of two overlapping spots: (a) initial image, (b) segmentation results obtained by the straightforward application of the Chan-Vese model.

Fig. 2. 3-D representations of protein spots: (a) partly overlapped and (b) highly overlapped.

Fig. 3. Multiple directions of straight-line segments for local intensity minima detection.

Fig. 4. (a) Real 2D-GE image, (b) detection results obtained by the local intensity minima process, (a₁) sub-image of (a), and (b₁) sub-image of (b).

Fig. 5. 2D-GE images obtained by the application of: (a) histogram equalization and (b) CLAHE, on the 2D-GE image of Fig. 4b. A sub-image of the original 2D-GE image is illustrated in (c), whereas the corresponding sub-images of (a) and (b) are magnified in 5(a₁) and 5(b₁), respectively.

Fig. 6. Histograms of: (a) the original 2D-GE image, (b) the image resulted from the application of histogram equalization on the image of Fig. 4b and (c) the image resulted from the application of CLAHE on Fig. 4b.

Fig. 7. Results obtained by the flood-fill morphological operation on: (a) the image illustrated in Fig. 4b and (b) on the enhanced image of Fig. 5b, which is generated by the application of CLAHE. A sub-image of the original 2D-GE image is illustrated in (c), whereas (a₁) and (b₁) are the corresponding sub-images of (a) and (b), respectively.

Fig. 8. 3-D representation of the level-set surface of multiple cones obtained by the application of the proposed initialization process on a real 2D-GE image.

Fig. 9. Segmentation results obtained by the application of: (a) the proposed scheme, (b) PDQuest 8.0.1, (c) Melanie 7, and (d) Delta2D software package, whereas (a₁)-(d₁) are sub-images of (a)-(d) respectively.

Fig. 10. (a) Synthetic 2D-GE image, and (b) the corresponding ground truth.

Fig. 11. Segmentation results of the application of: (a) the proposed scheme, (b) PDQuest 8.0.1, (c) Melanie 7 and (d) Delta2D software package, (a₁)-(d₂) sub-images of (a)-(d) respectively.

Fig. 12. Overall segmentation results in terms of v_o and v_e , obtained by the proposed scheme, as well as by PDQuest 8.0.1, Melanie 7 and Delta2D software packages, on the set of synthetic 2D-GE images.

Figures

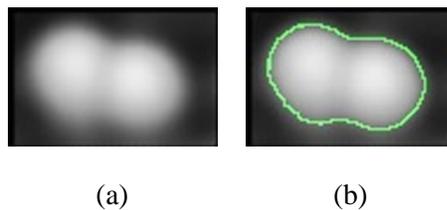
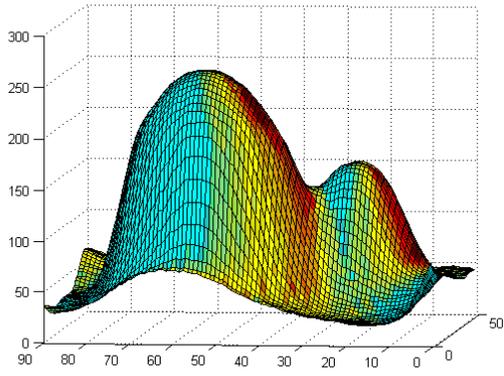
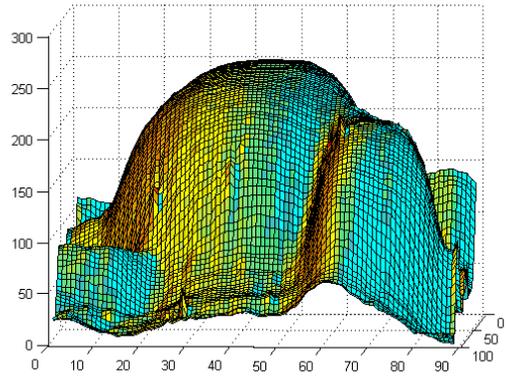


Fig. 1



(a)



(b)

Fig. 2

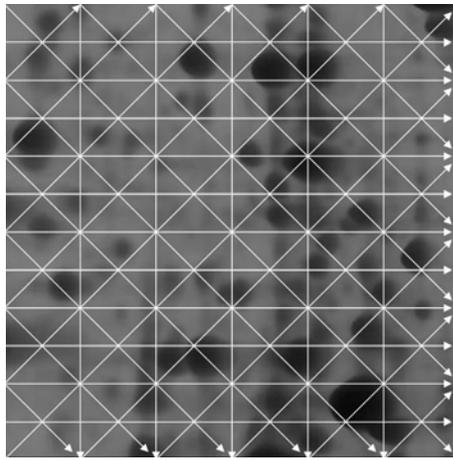
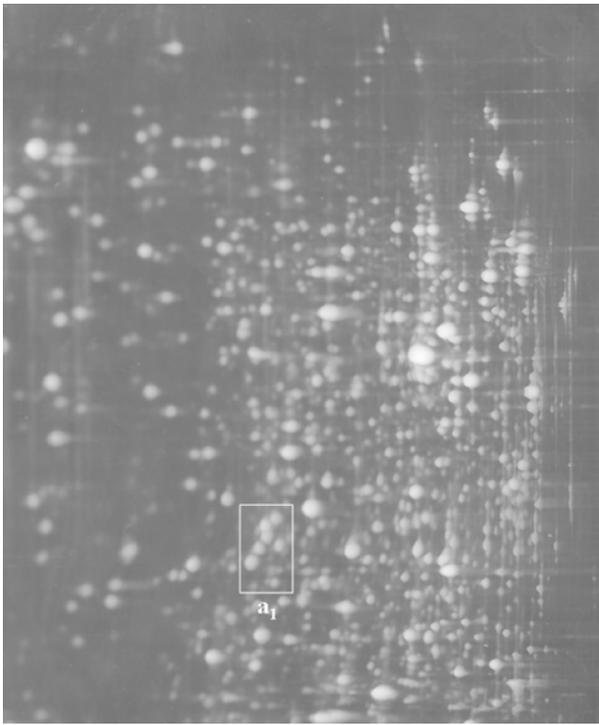
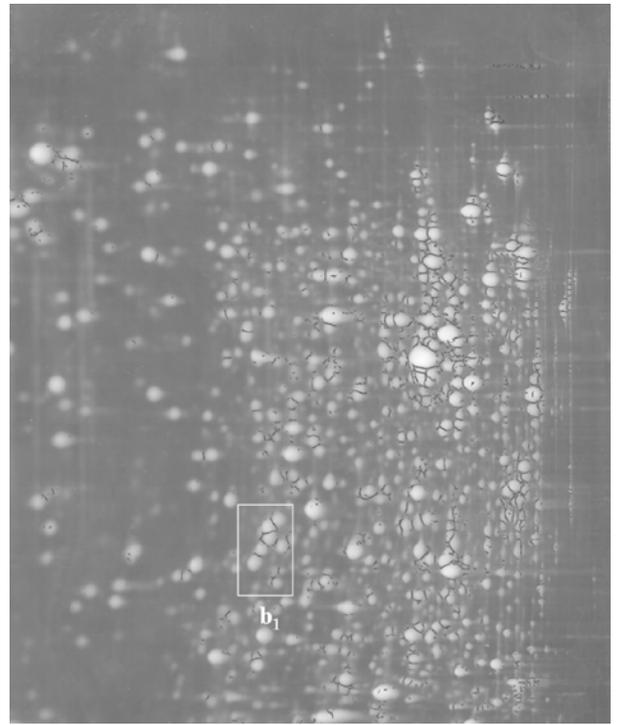


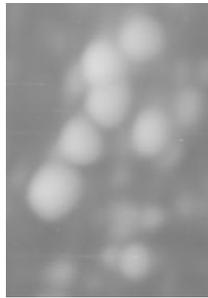
Fig. 3



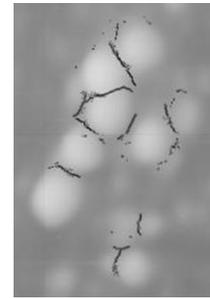
(a)



(b)

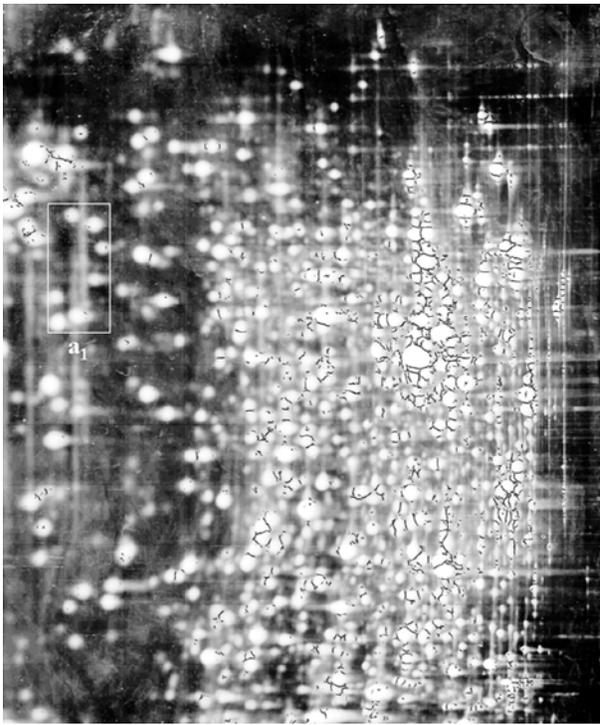


(a₁)

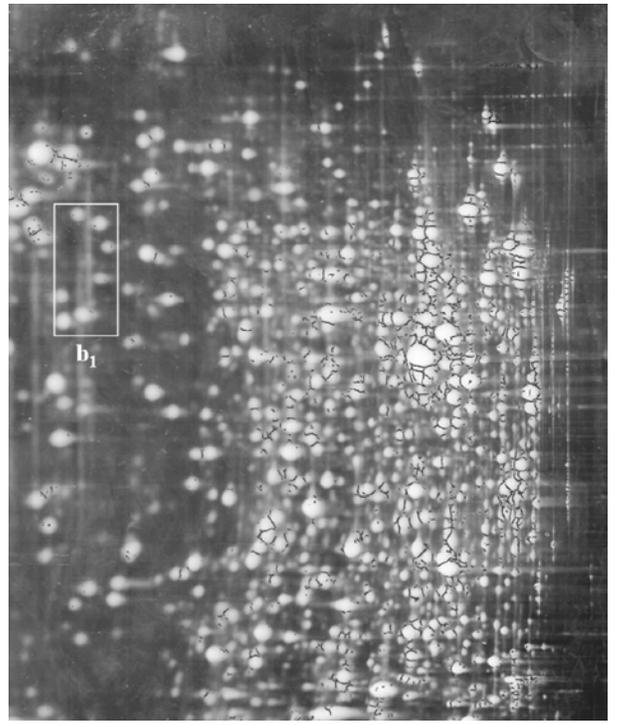


(b₁)

Fig. 4



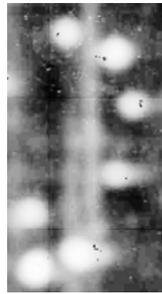
(a)



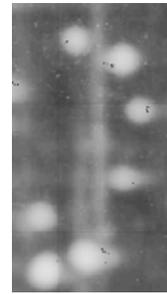
(b)



(c)

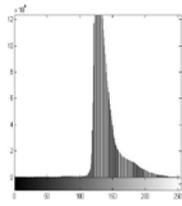


(a₁)

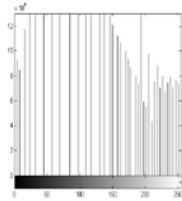


(b₁)

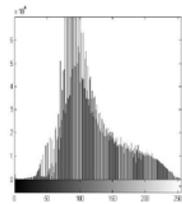
Fig. 5



(a)

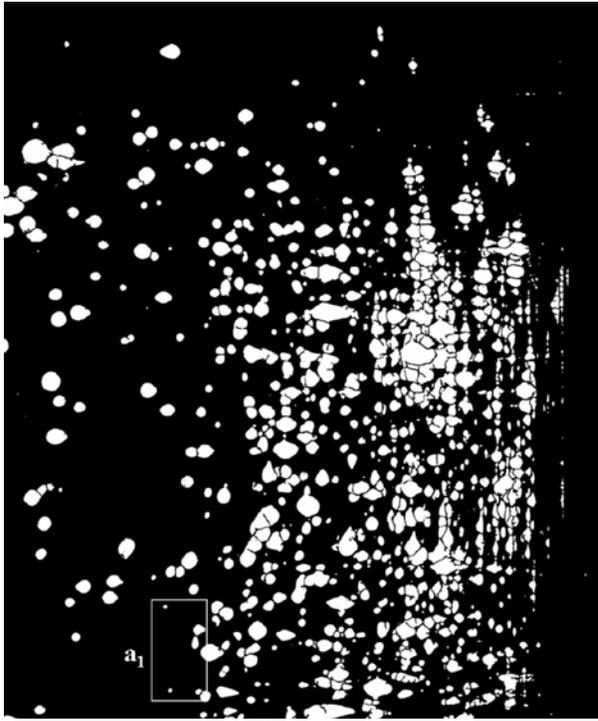


(b)

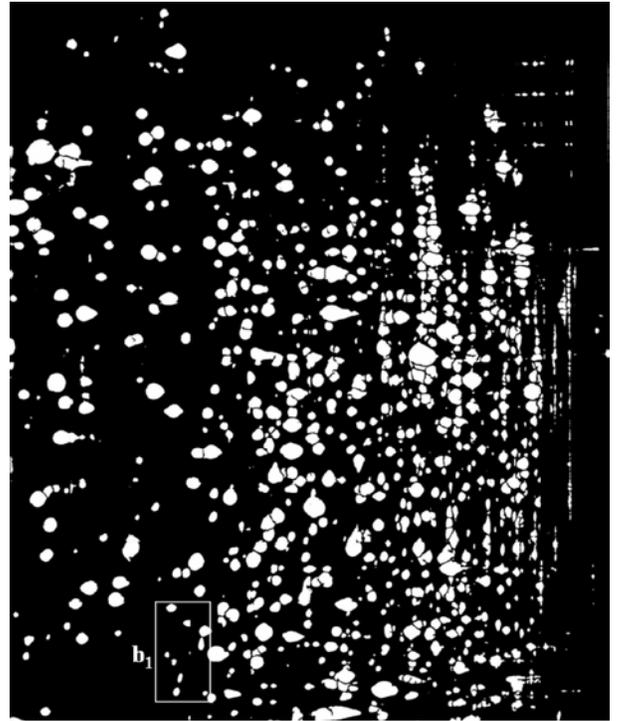


(c)

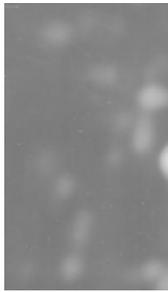
Fig. 6



(a)



(b)



(c)



(a₁)



(b₁)

Fig. 7

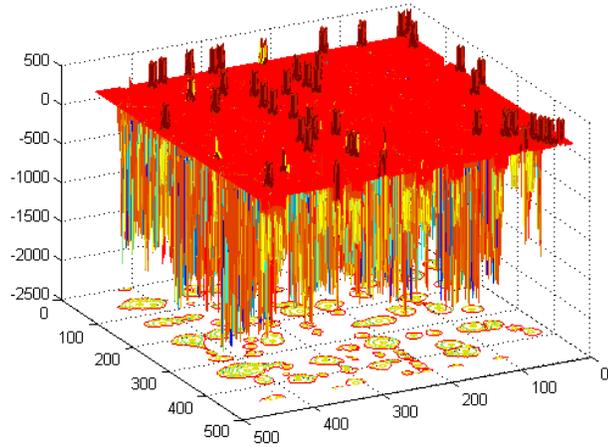
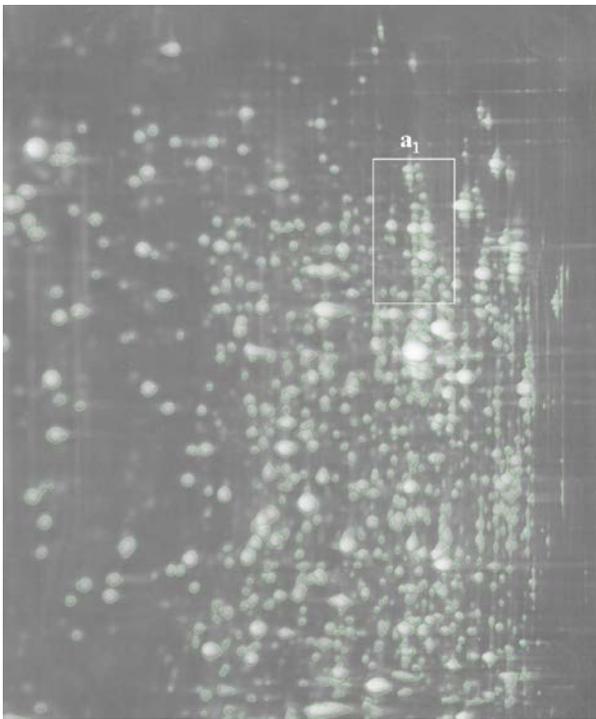
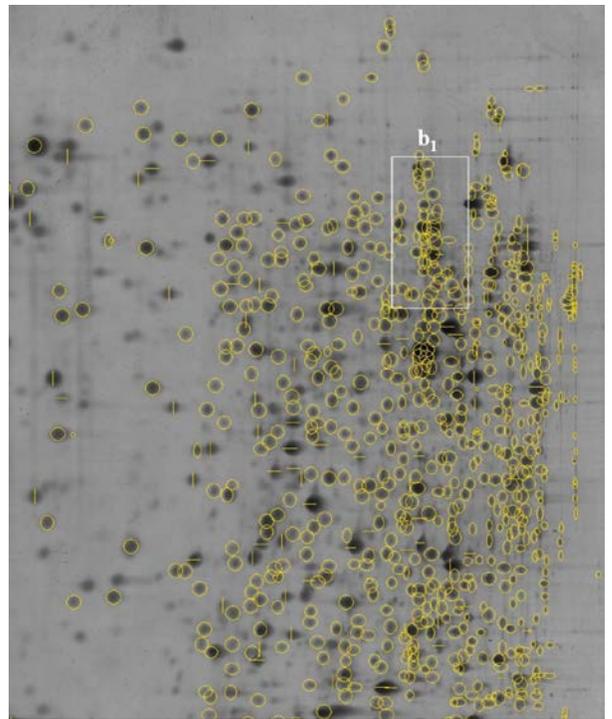


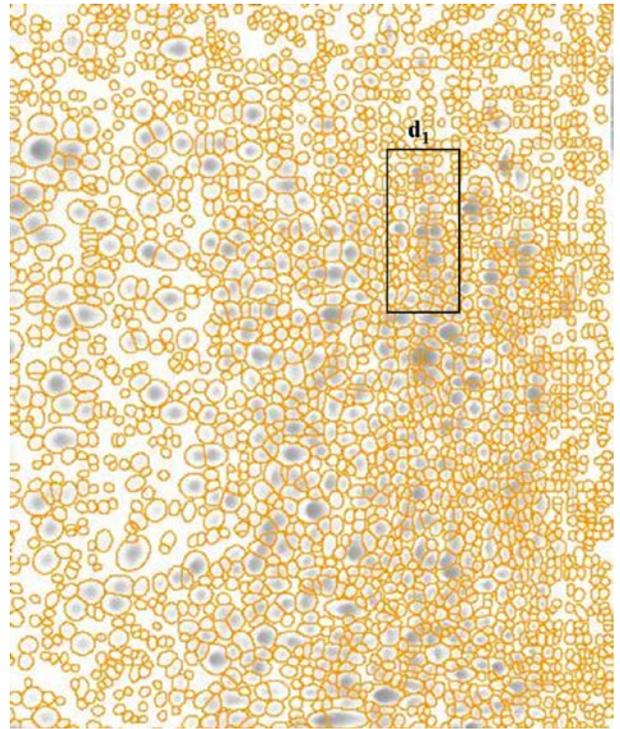
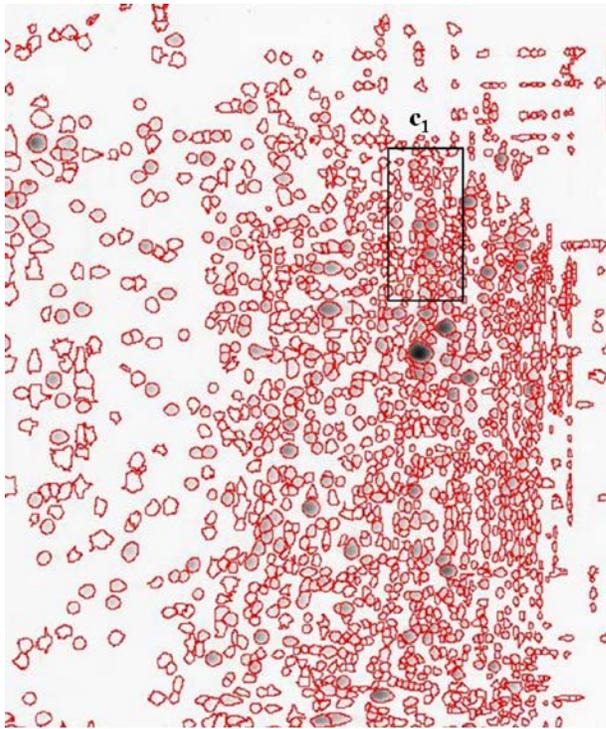
Fig. 8



(a)

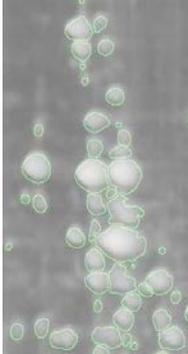


(b)

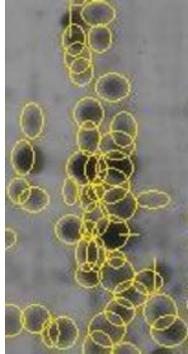


(c)

(d)



(a₁)



(b₁)

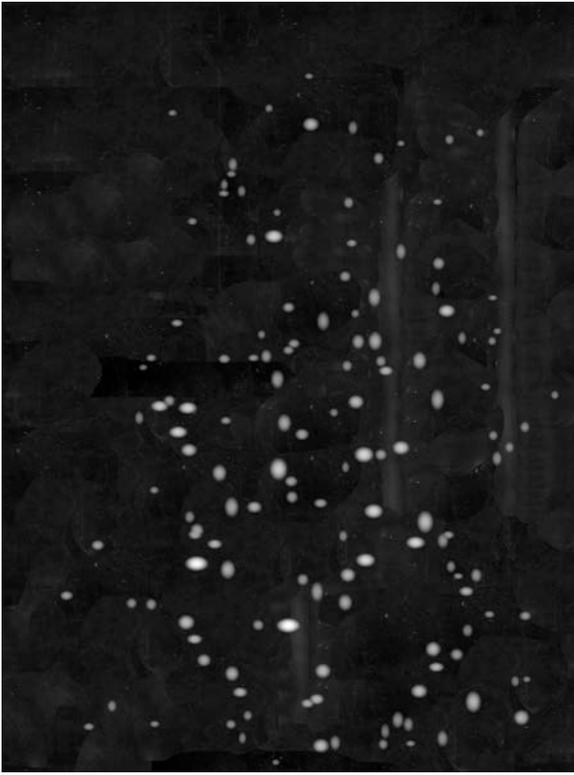


(c₁)

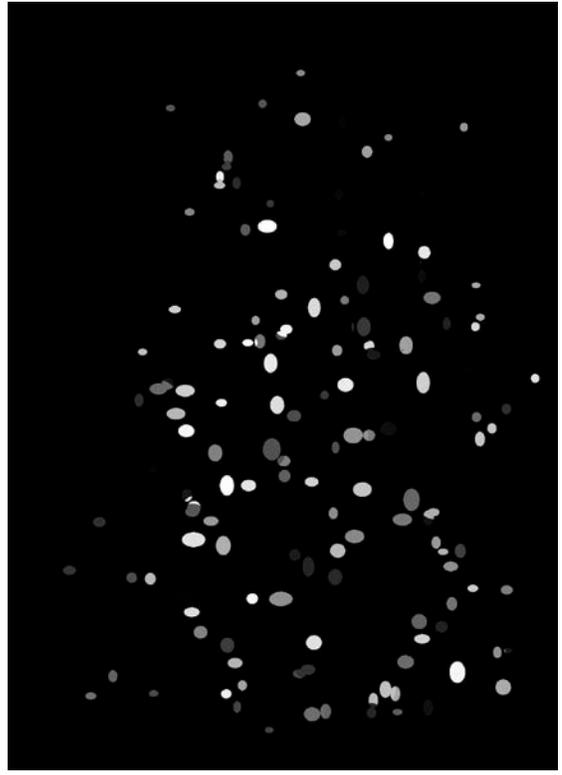


(d₁)

Fig. 9

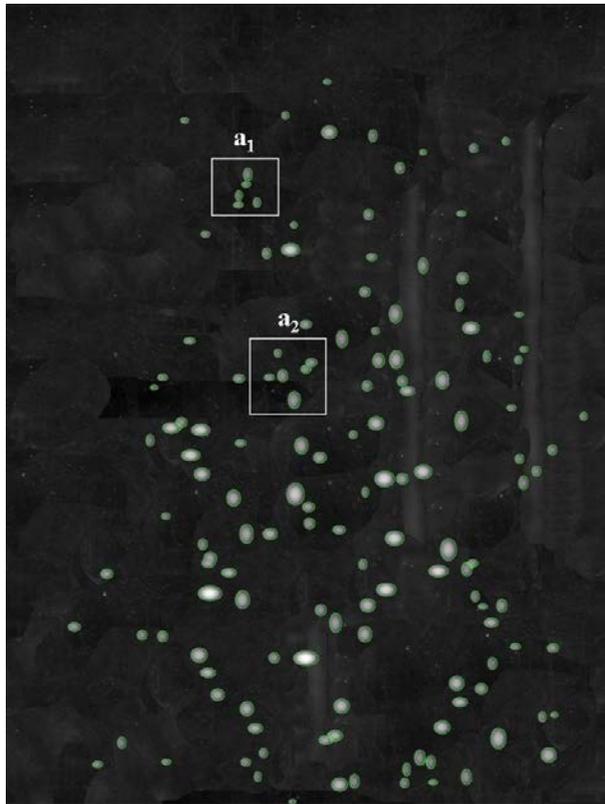


(a)

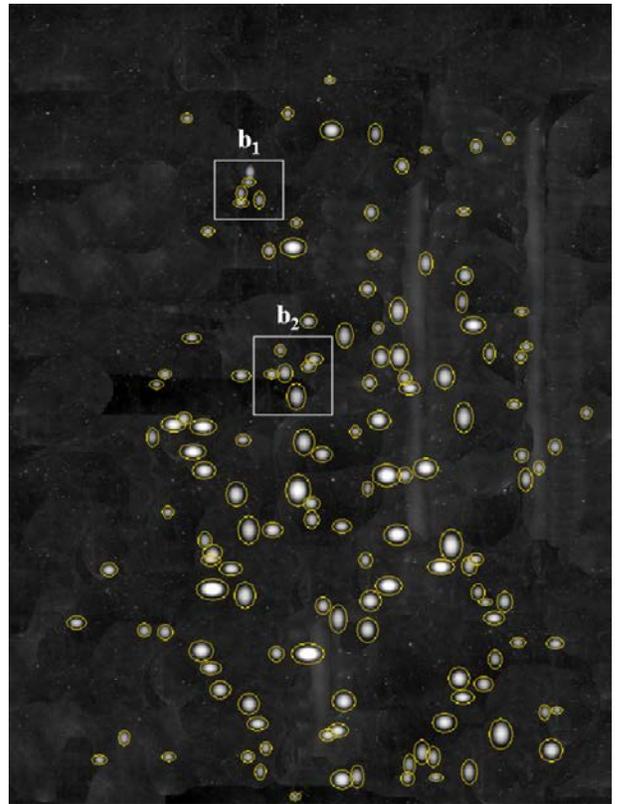


(b)

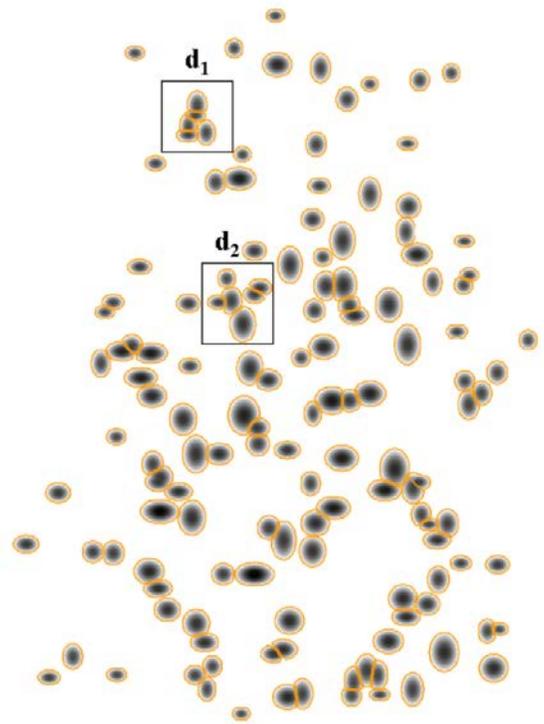
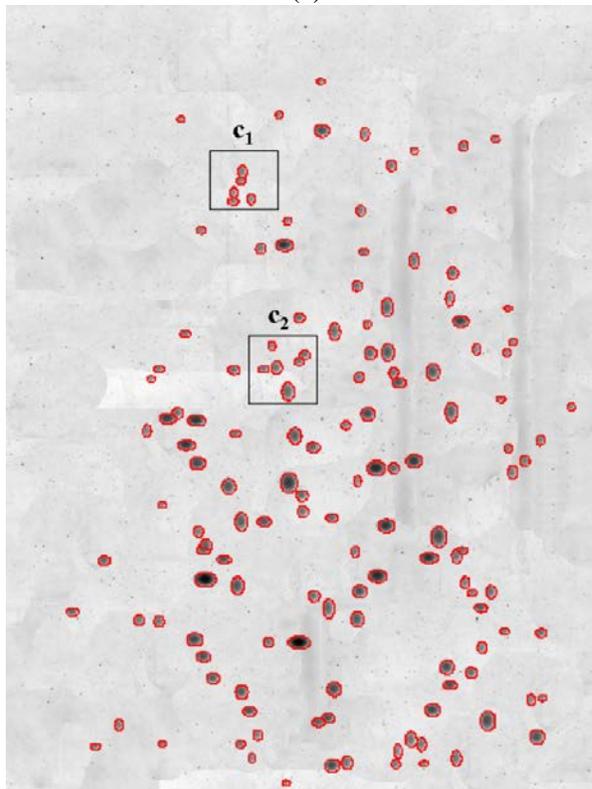
Fig. 10



(a)



(b)



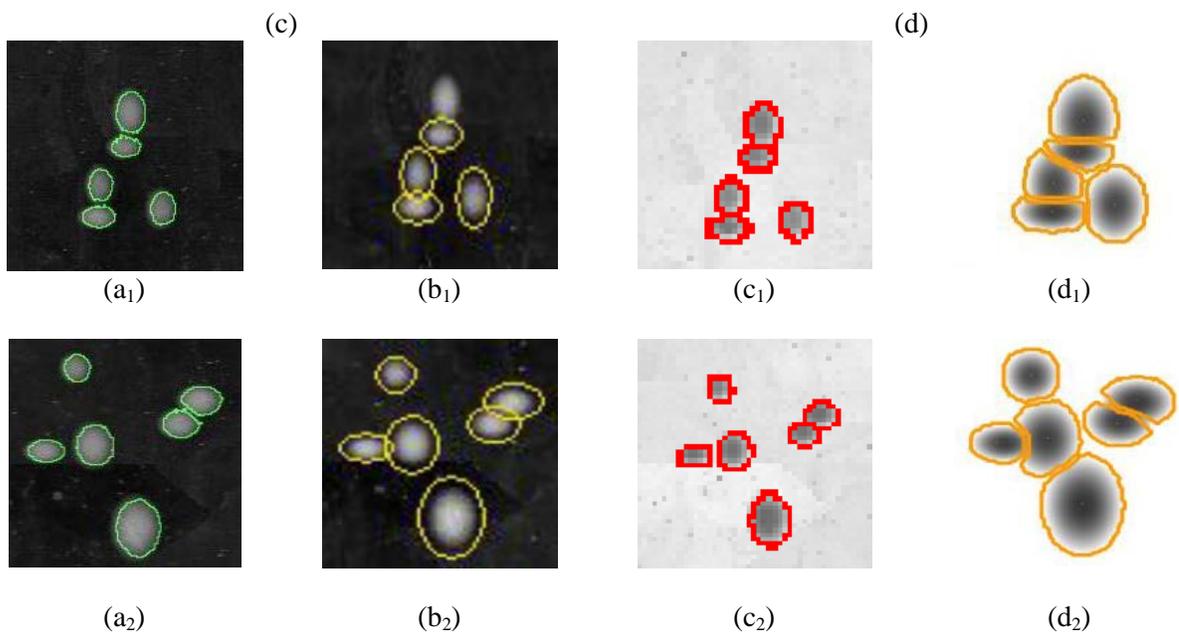


Fig. 11

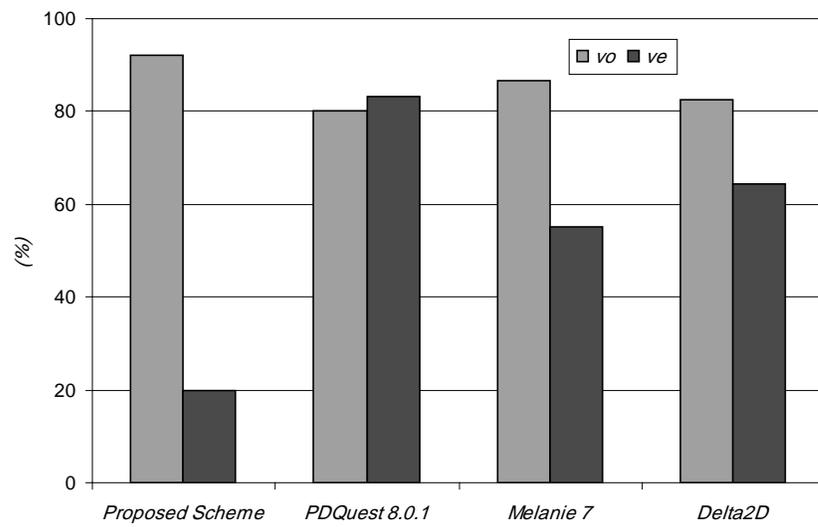


Fig. 12