Machine learning methods for effective proteomics image analysis

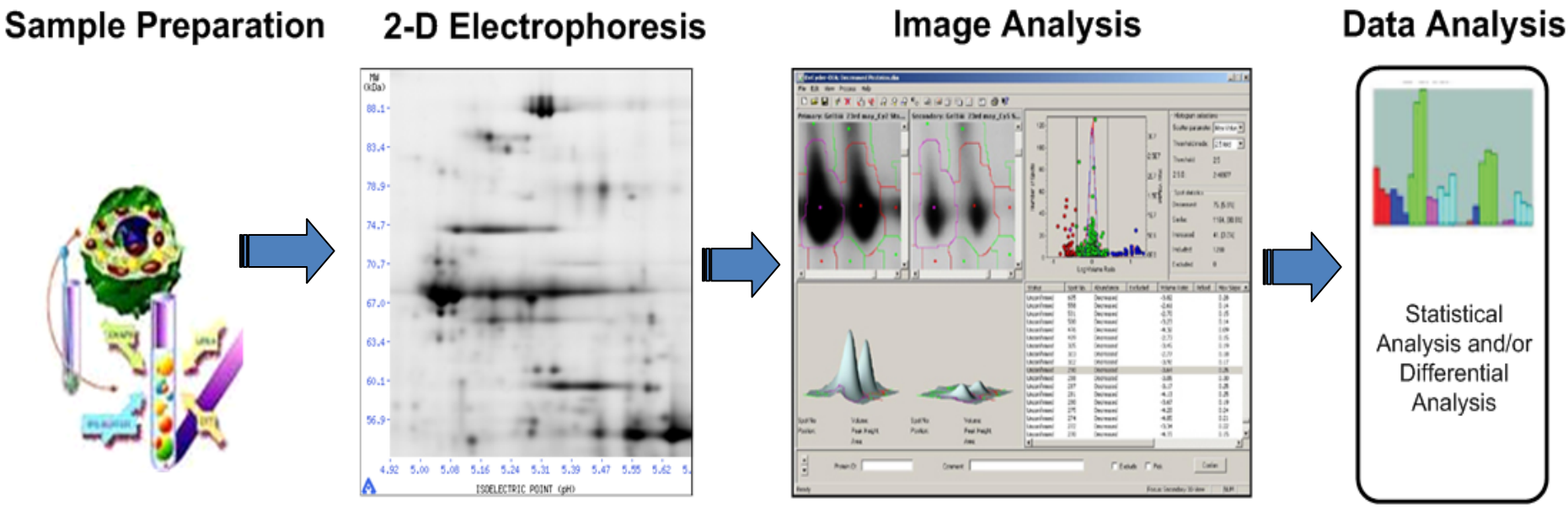
Panos Tsakanikas and Elias S. Manolakos

Dept. of Informatics and Telecommunications
University of Athens, Greece

Overview

- Proteomics analysis workflow (2DGE)
- Research motivation
- End-to-end image analysis pipeline
  1. 2DGE image denoising using the Contourlets Transform
  2. Hierarchical image segmentation
     a. Regions of Interest extraction before spot detection
     b. Protein spots detection and volume estimation using machine learning methods
- Conclusions
- Future research directions

2DGE image analysis

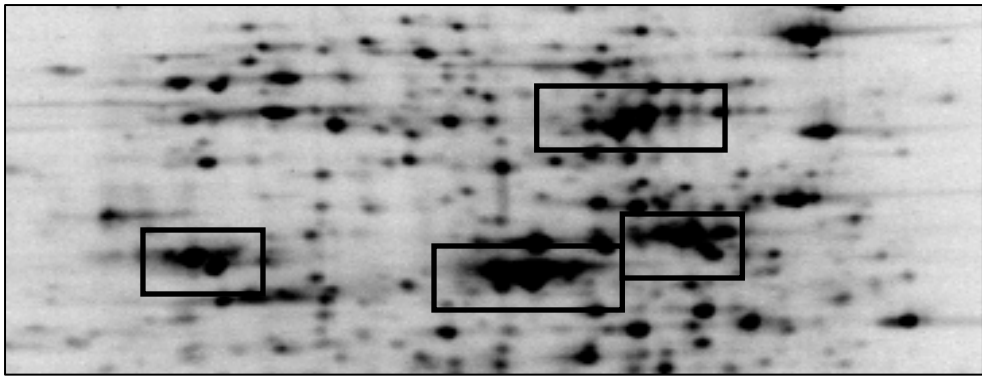


Differential proteomics data analysis facilitates:

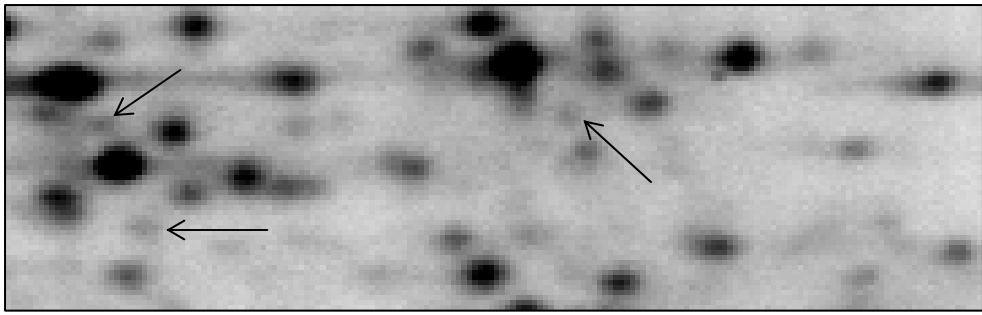
Biomarkers discovery for personalized drug design

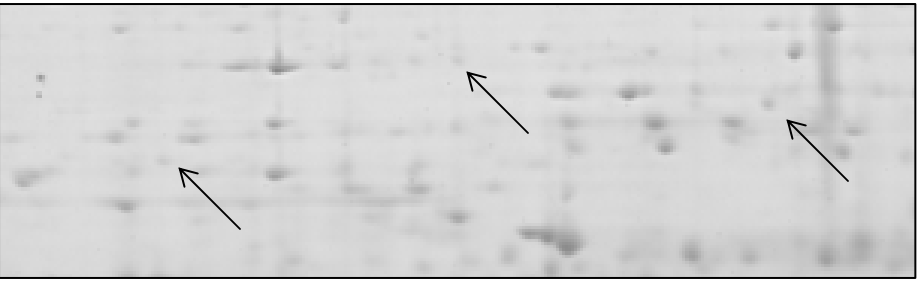
Difficulties with 2DGE image analysis

- Areas with overlapping and/or saturates protein spots



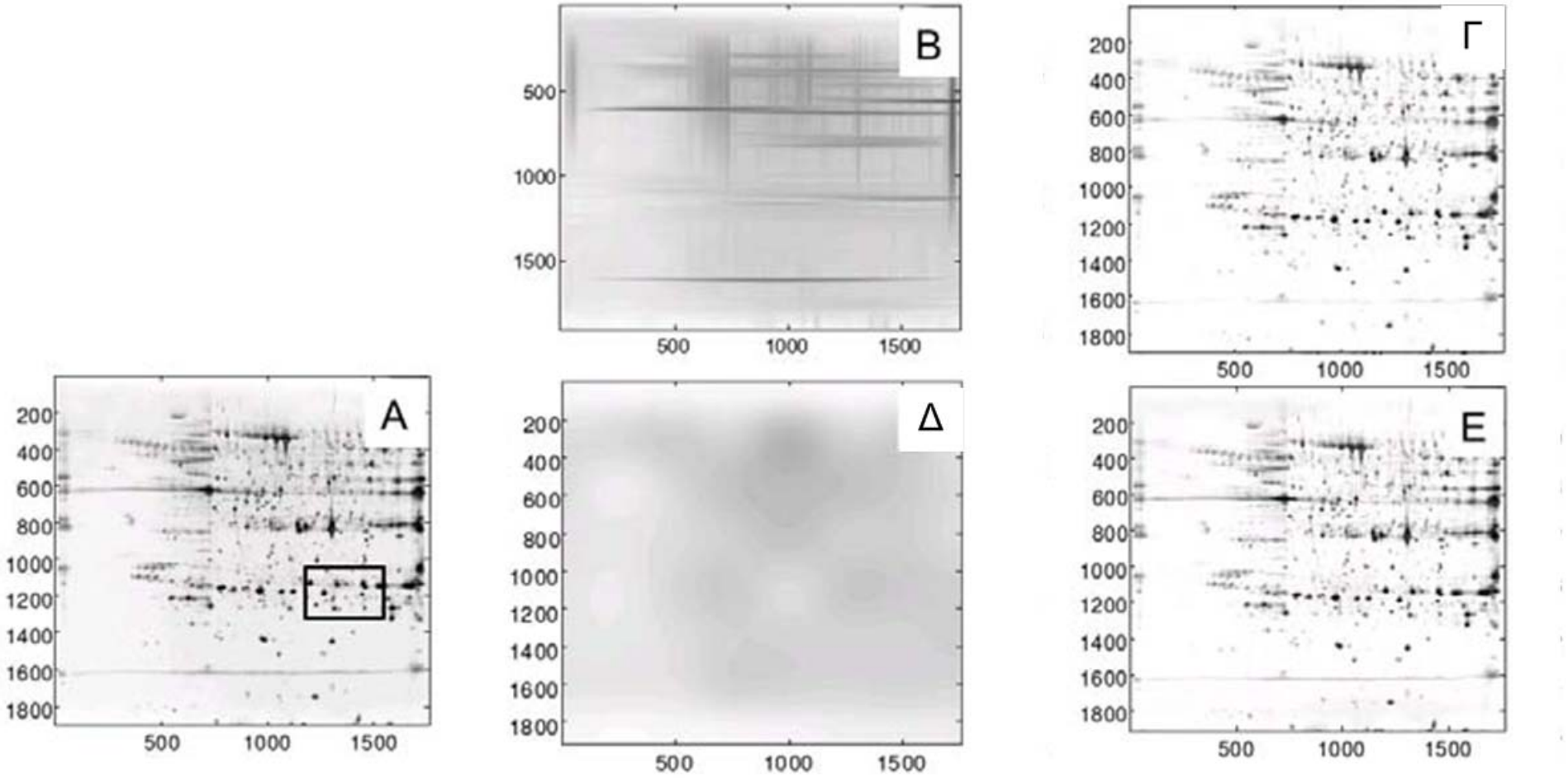
- Spots of low intensity (faint) that cannot be detected
- Small faint spots next to very big bright spots





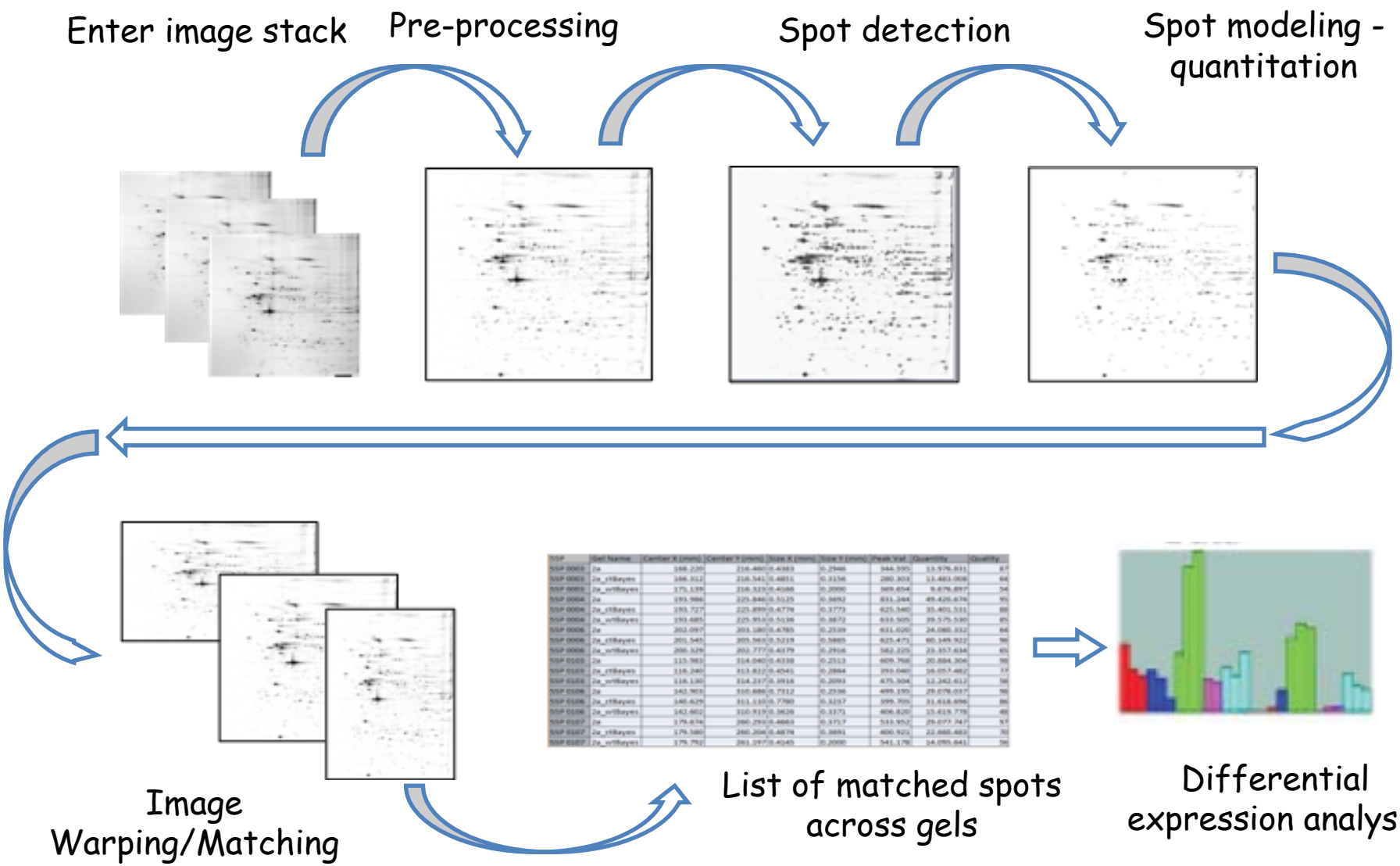
Difficulties with 2DGE image analysis

Additive background



Multiplicative background

2DGE differential proteomics workflow



Research Motivation

The myth of high throughput proteomics image analysis using commercial software packages

- ✓ Different packages give very different spot detection results for the same image
- ✓ Substantial manual spot editing is required to correct software errors
- ✓ Less than 3% of the total image analysis time corresponds to automated processing!

Commercial Package	Mean number of detected spots	Mean after manual spot editing	Time (in hours) to process 18 images
Delta 2D	820 ± 0	846 ± 0	1
Imagemaster	568 ± 134	626 ± 68	22
PDQuest	1395 ± 639	703 ± 35	10
Progenesis	1471 ± 268	674 ± 32	18
ProteinMine	893 ± 380	706 ± 60	18

Commercial software packages are useful, but
- are far from automated,
- require tuning parameters for every new image

Clark, Br. N., Gutstein H. B., "The myth of automated,high-throughput 2D gel analysis", *Proteomics 2008.*

Summary of Contributions

- We have developed an and-to-end fully automated image analysis pipeline:

Stage1: 2DGE image denoising using the Contourlets Transform

Tsakanikas, Manolakos, [Proteomics 2009]

Stage2: Hierarchical image segmentation

a. Isolate Regions of Interest (ROI) from the image

Tsakaniakas , Manolakos [EUSIPCO 2008]

b. Detect and quantify spots inside each ROI using machine learning methods

Tsakaniakas , Manolakos [Proteomics2010 under review]

Traditional image segmentation approaches

All known spot detection methods attempt to extract **in one step** the individual spots, something that presents the following problems:

- ✘Small-size faint spots may get lost in the background
- ✘If we try to avoid this, a lot of extraneous spots are introduced
- ✘Areas with complex overlapping spots may not be segmented properly

The most popular spot segmentation method is using the Watershed Transform, but is known to lead to image over-segmentation

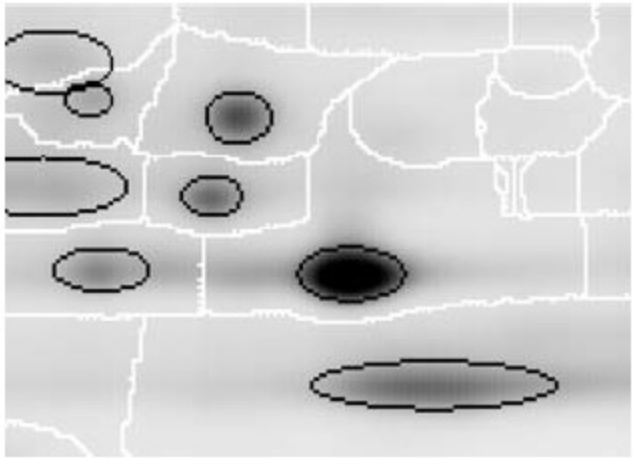


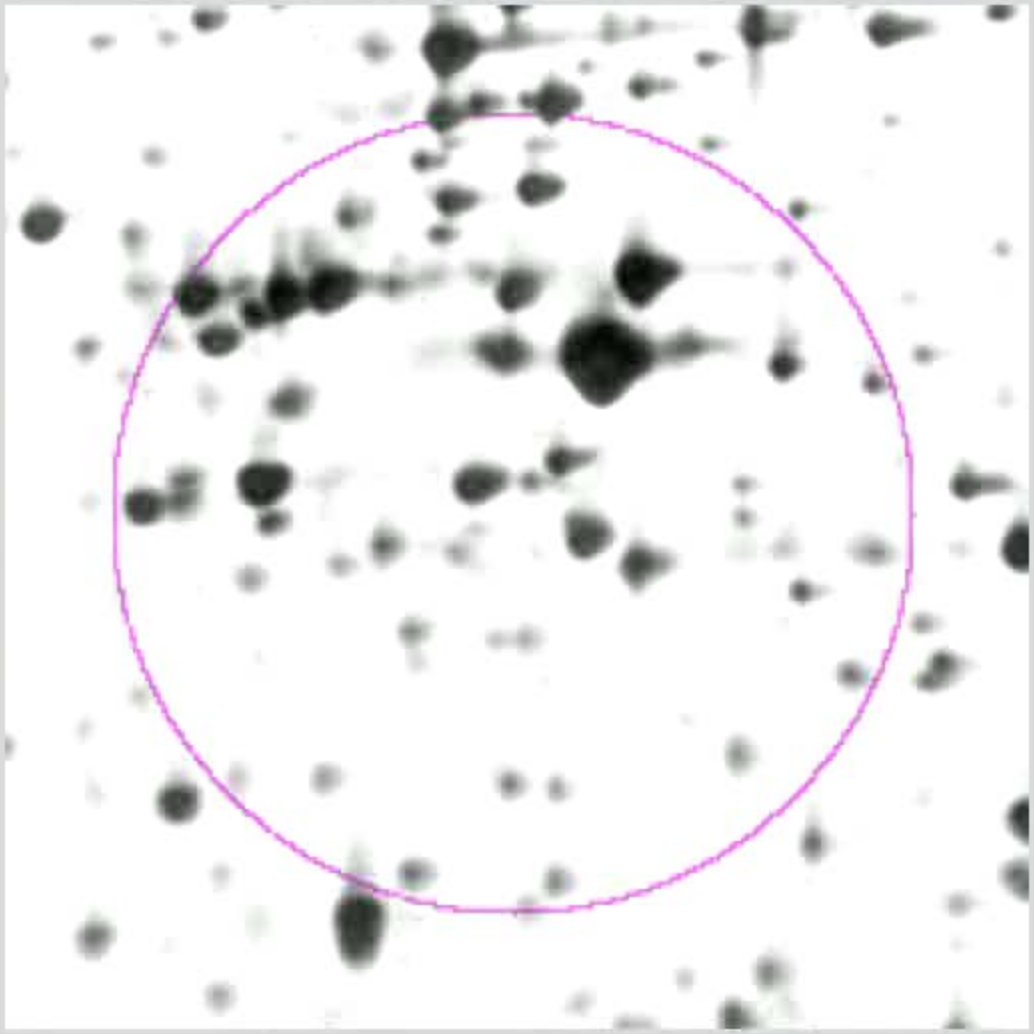
Image spot detected using the Watershed transform and Gaussian spot modeling in each segmented region

Observe the oversegmentation and the fact that extracted areas are not "tight" i.e. still include a lot of background pixels

We introduced a **2-step** spot detection approach:

1. Separate the **Regions of Interest** (which may contain spots) from the background areas (without true spots) using Active Contours
2. Analyze each ROI using **machine learning** methods to "fish out" the individual spots

ROIs extraction algorithm - Demo



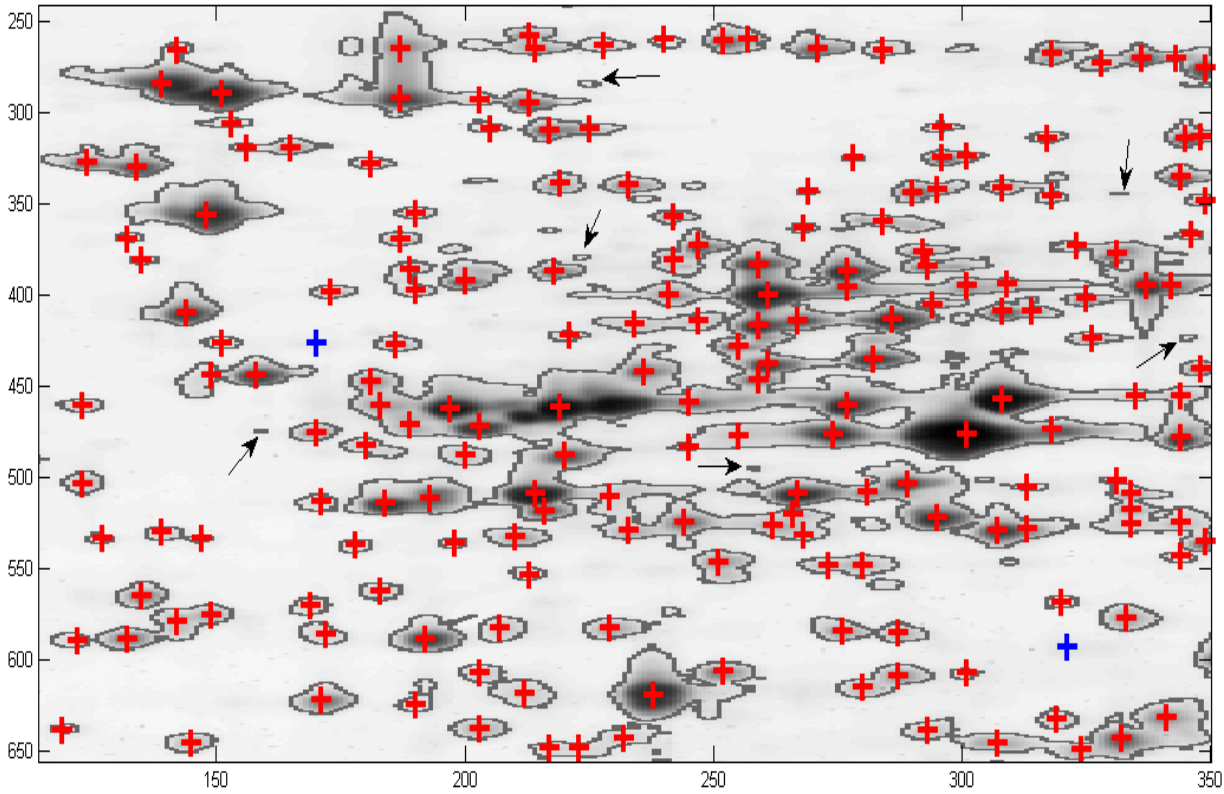
ROIs extraction algorithm - Validation

- ✓Red crosses → True spots inside extracted ROIs by our method and detected by PDQuest
- ✓Blue crosses → True spots in missed ROIs, detected by PDQuest
- ✓ROIs pointed by small arrows → False Positive ROIs we have introduced
- ✓ROIs without crosses → ROIs with True spots that PDQuest has missed

Image	TPF	PPV
1a	99,55%	99,55%
2a	99,30%	99,61%
MP1	99,23%	96,28%
MP2	94,68%	98,42%
MP3	99,59%	97,57%
Rj1	91,97%	99,21%
RGA	97,88%	99,35%
RGB	98,86%	98,77%

$$TPF = \frac{True\ Positives}{True}$$

$$PPV = \frac{True\ Positives}{Positives}$$



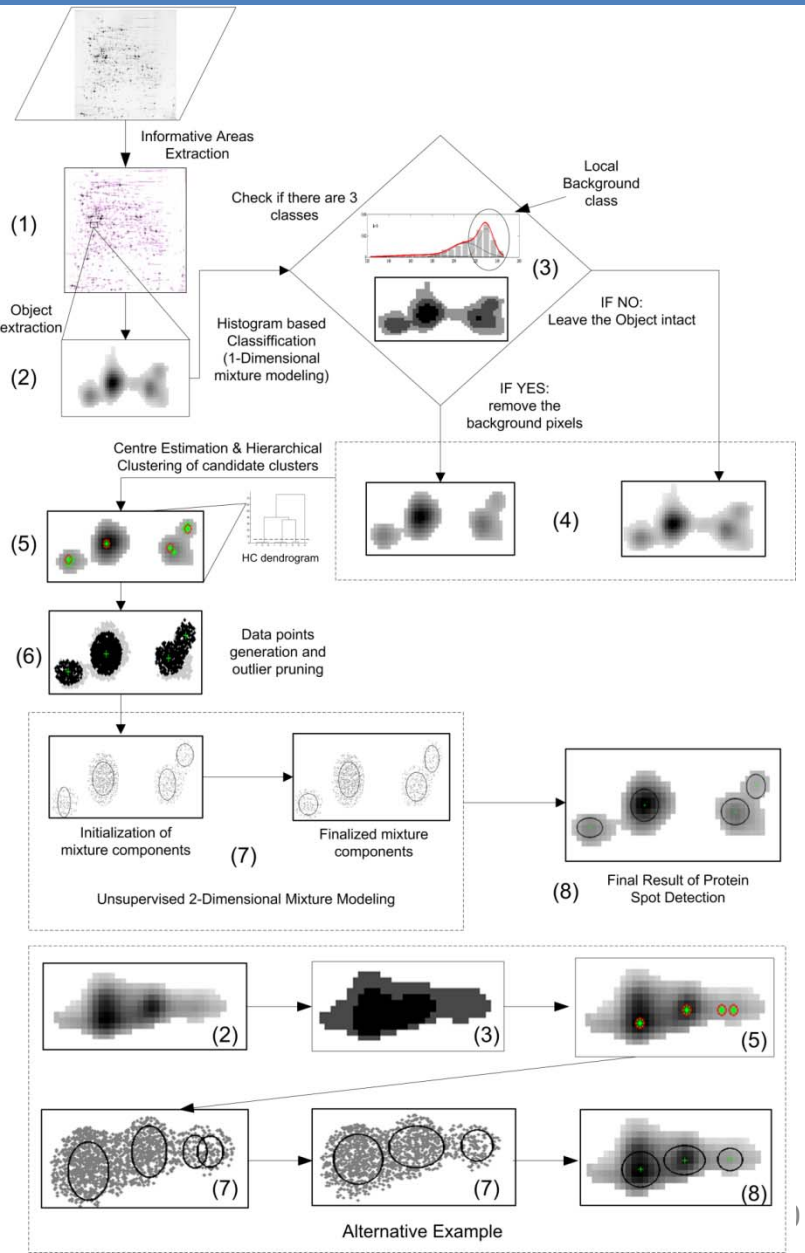
Evaluation using a lot of synthetic and real images:

The very large majority of true spots are confined inside the extracted ROIs

Fishing for spots inside ROIs - overview

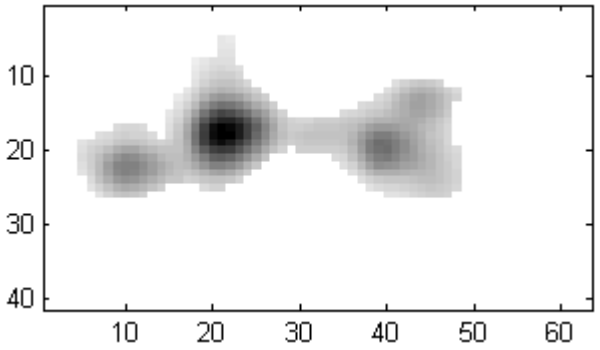
Our Spot detection approach:

1. Extract Regions of Interest
2. Remove remaining background pixels
3. Initial spot centers estimation
4. Reverse engineering: From pixels to underlying distributions of protein species molecules
5. Remove outliers
6. Spot modeling using 2D Gaussian mixtures
7. Spot center and quantity estimation (GEM with model selection)



Fishing for spots - background pixels removal

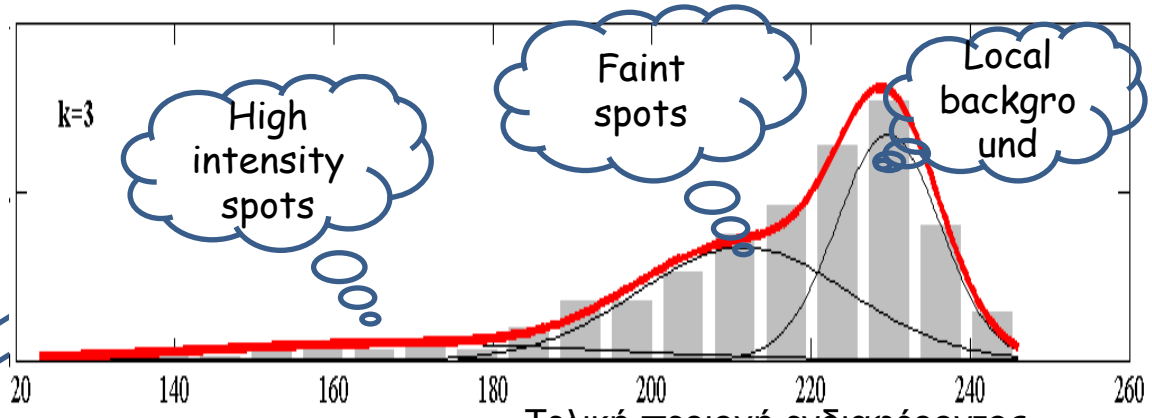
Extracted image object(ROI)



Local background removal:

Fitting 1D Gaussian models to the pixels intensities histogram

Three classes of pixels are expected at most (normal spots, faint spots , background/streaks)



Classification results

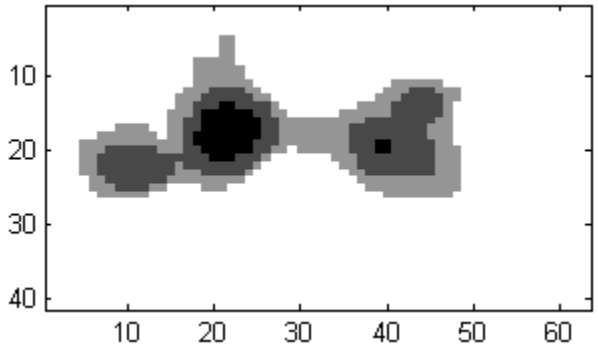
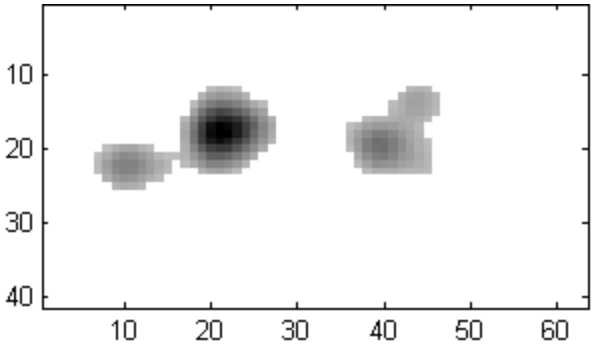


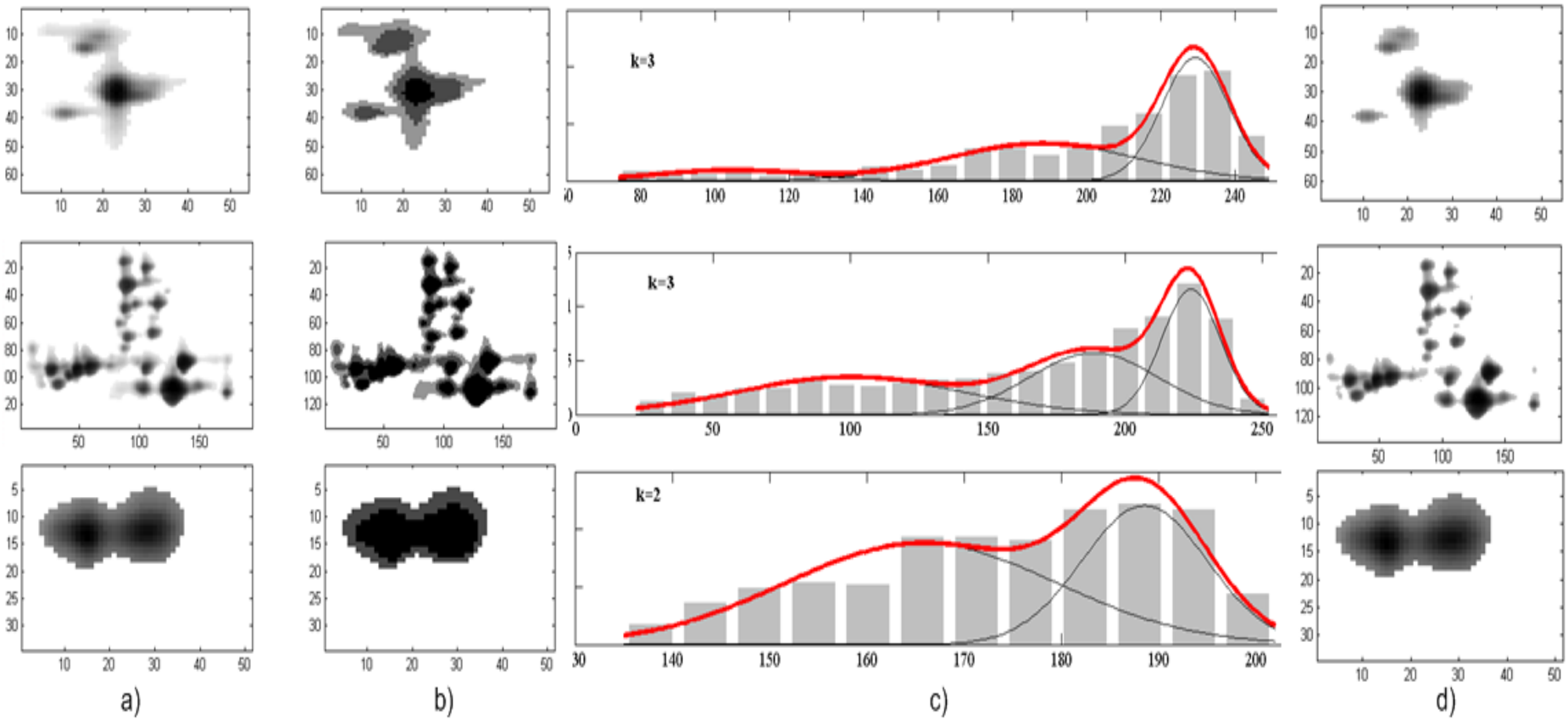
Image object after removing local background

Τελική περιοχή ενδιαφέροντος



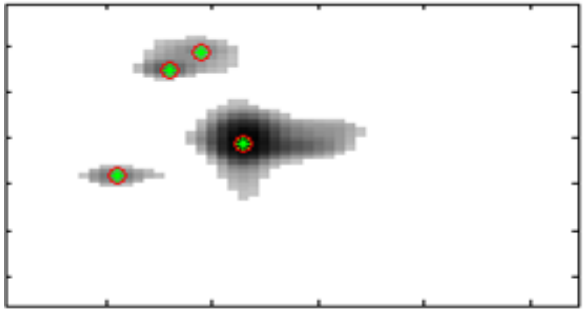
Local Background removal - examples

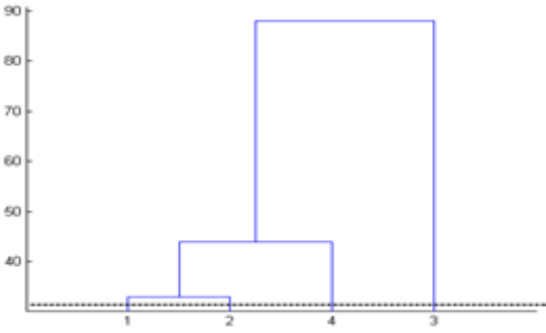
Use 1D Gaussian models with up to κ=3 components.
Select the best mixture model, among those with κ=1, 2, 3, using the Minimum Message Length (MML) criterion.
Use the EM algorithm to fit the mixture models

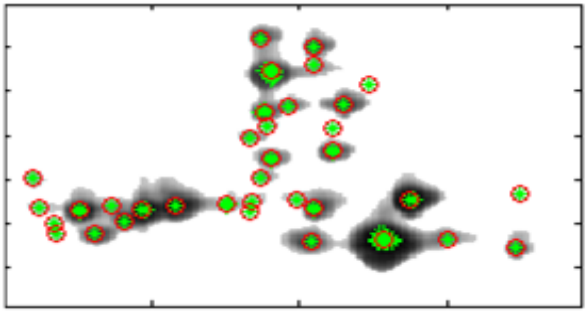


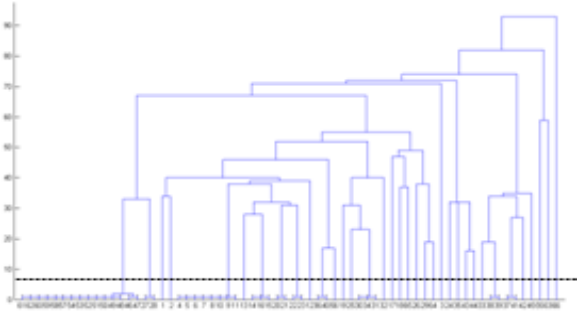
Fishing for spots – locating candidate spot centers

- Apply a 5x5 spatial filter to identify all the local intensity minima (green points) in the image object.

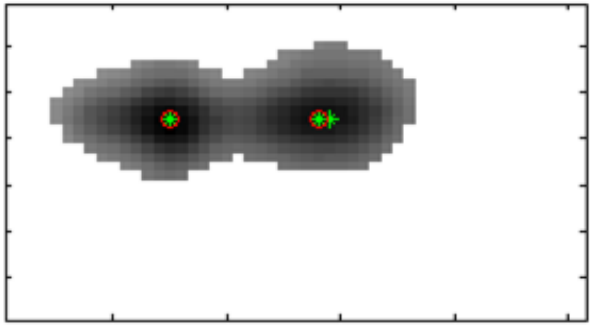


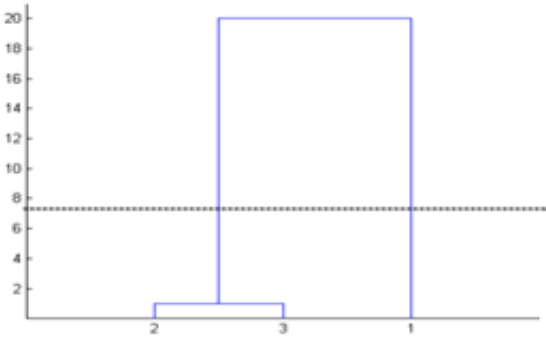






- Apply Hierarchical Clustering By cutting the dendrogram at a certain height we can estimate candidate spot centers (red circles).





Inverse modeling – from pixels to molecules

Reverse engineering approach:

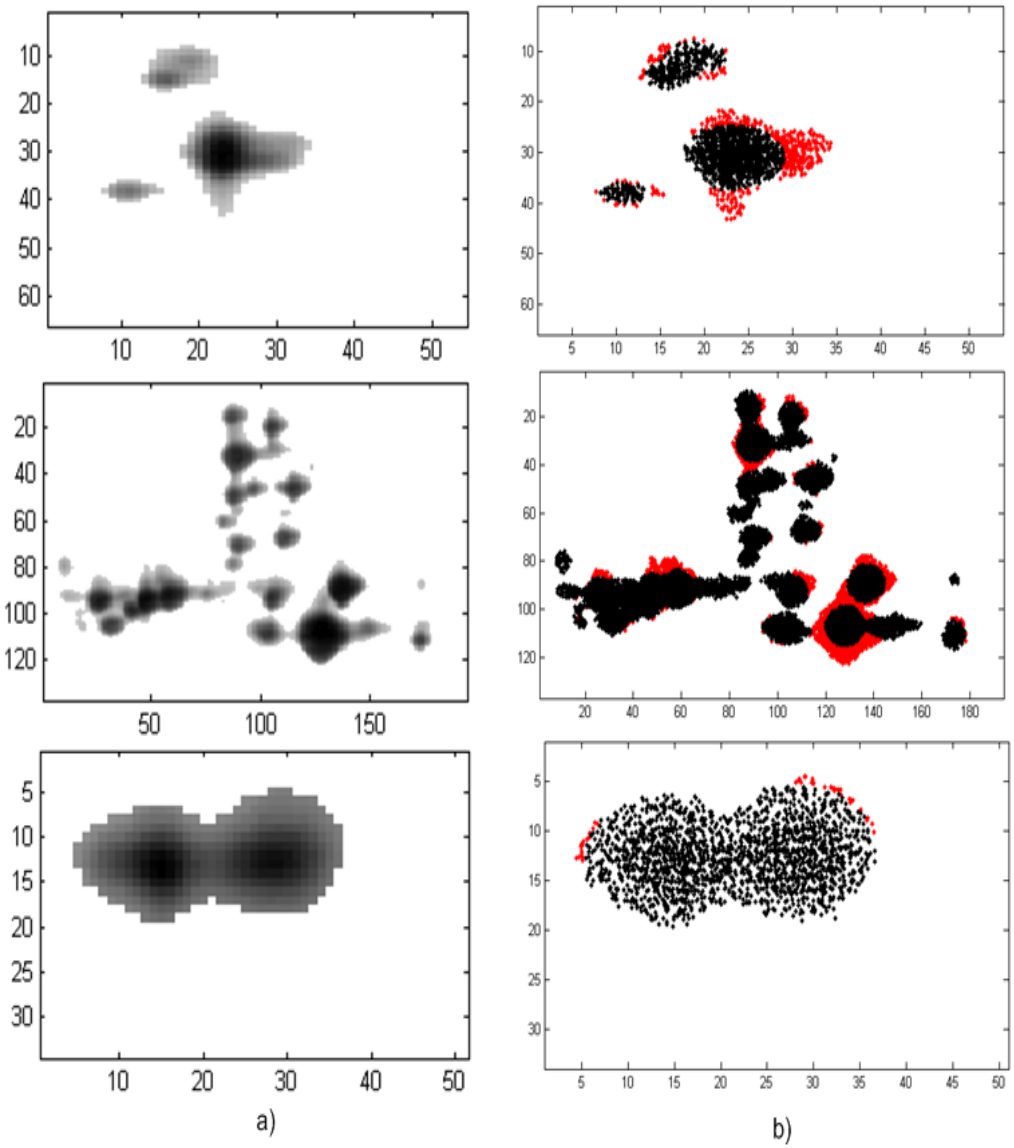
✓Each pixel represents a certain concentration of protein molecules at a particular location of the gel .

The image object can be thought as a set of points generated by

- ✓ a mixture of as many models as the number of pixels,
- ✓ with mixing coefficients proportional to the relative pixel intensities: $\pi(i) = I_i / \sum_{j=1}^{M} I_j$

Molecules dataset generation mechanism:

Distribute a total of C*N points ("molecules") to the pixels of each image object, according to the mixing coefficients π(i).
Treat each pixel as a molecules generator that "throws" those points randomly around its location.



Spot detection - Finding the best mixtures model that fits the generated dataset

Apply EM algorithm:

- Initialization:
  - Perform 1-NN classification
- Initial component centers
  - Estimated spot centers by HC
- Initial mixture coefficients
  - Percentage of points belonging to each initial center after 1-NN

Initial covariance matrices

- Sample covariance matrices after 1NN classification

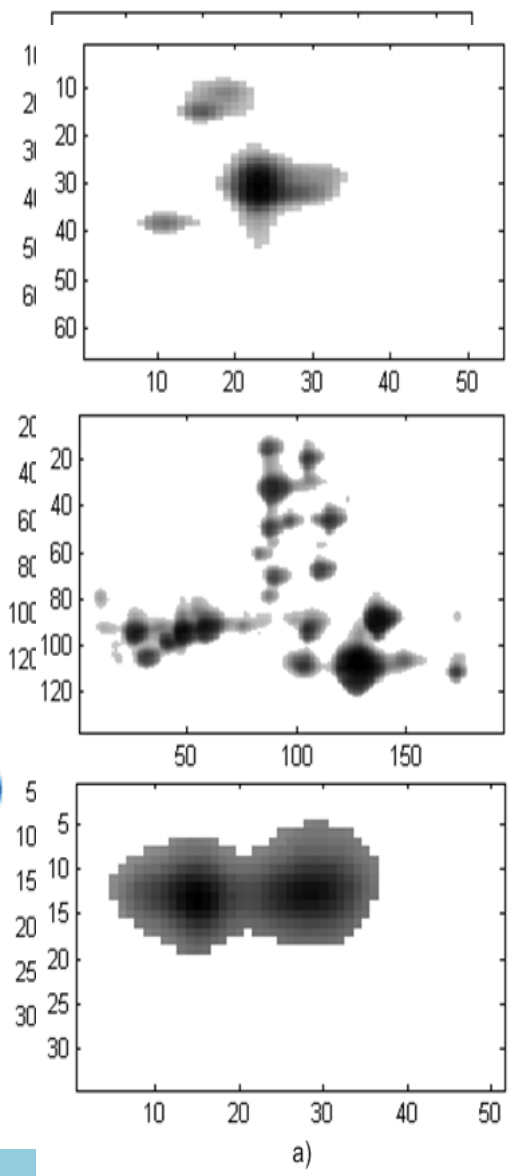
- Outliers pruning

$$\log p(\boldsymbol{y}^{(i)}|\boldsymbol{\theta}) = log \sum_{m=1}^{C} w_m p(\boldsymbol{y}^{(i)}|\boldsymbol{\theta}_m)$$

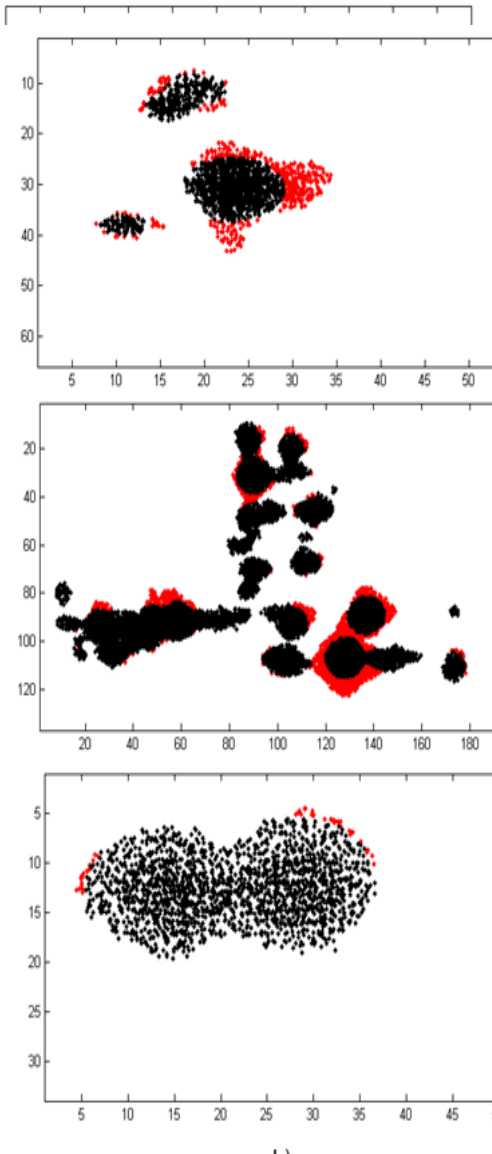
$$\log p(\boldsymbol{y}^{(i)}|\boldsymbol{\theta}) < 0.1 * max_{m=1:C}(\log p(\boldsymbol{y}^{(i)}|\boldsymbol{\theta}_m))$$

Use the Minimum Message Length criterion (MML) to identify the number of components giving the best results.

It is possible to reject initial spot centers!

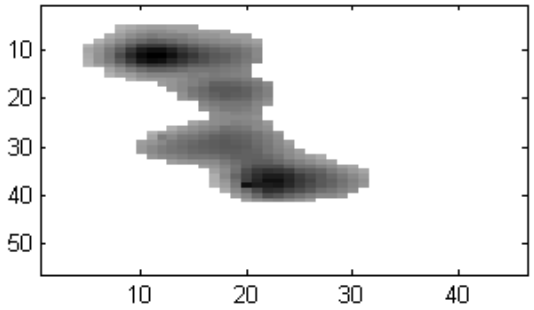


a)

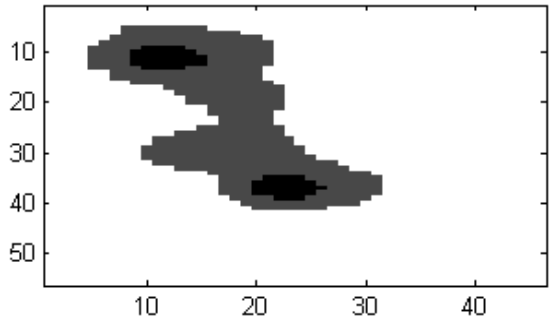


b)

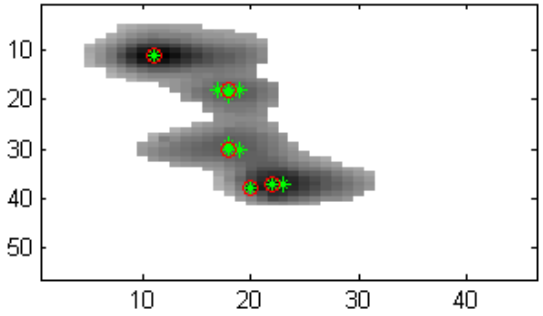
Examples of applying the proposed method



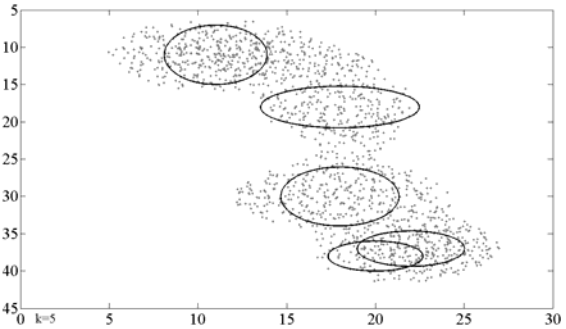
Initial image object (ROI)



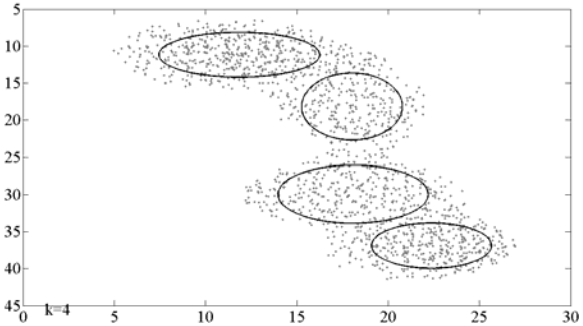
Pixels classification
(2 classes, no background pixels need to be removed)



Spot centers estimation
(5 candidate centers)



Initial spot modeling
(with 5 components)



Final result after the EM
(4 components give the best result)

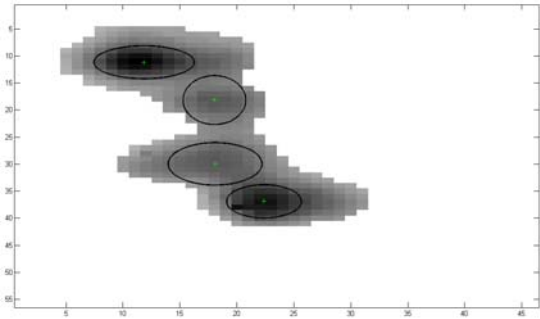
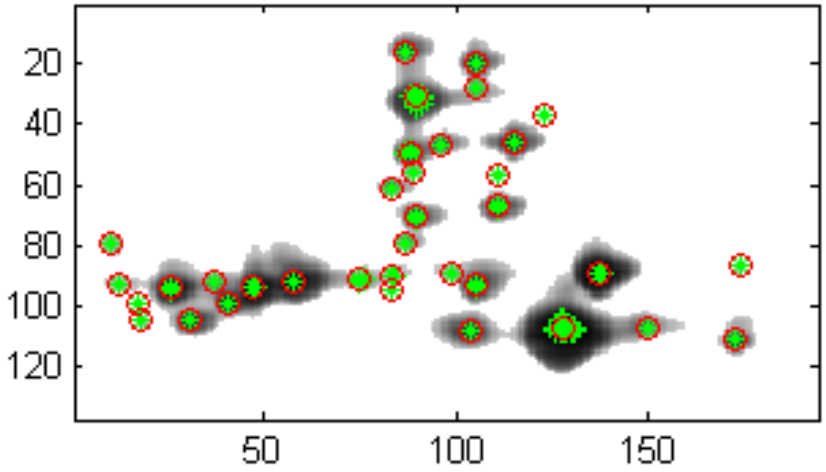


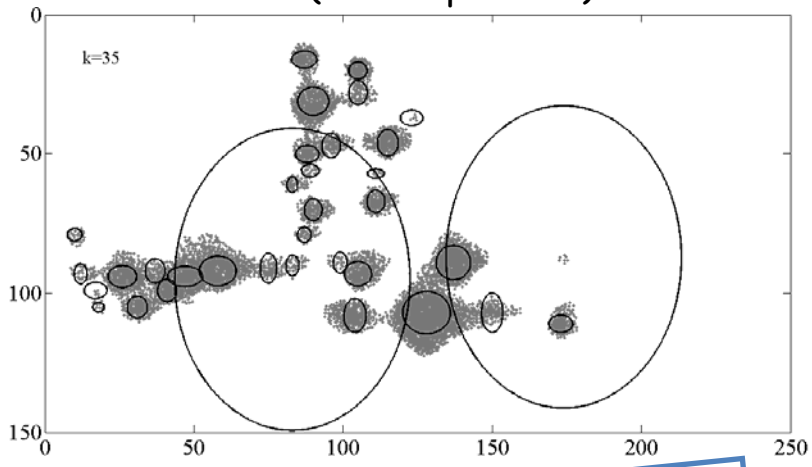
Image object after spot detection
(4 spots detected)

More complex spot detection example

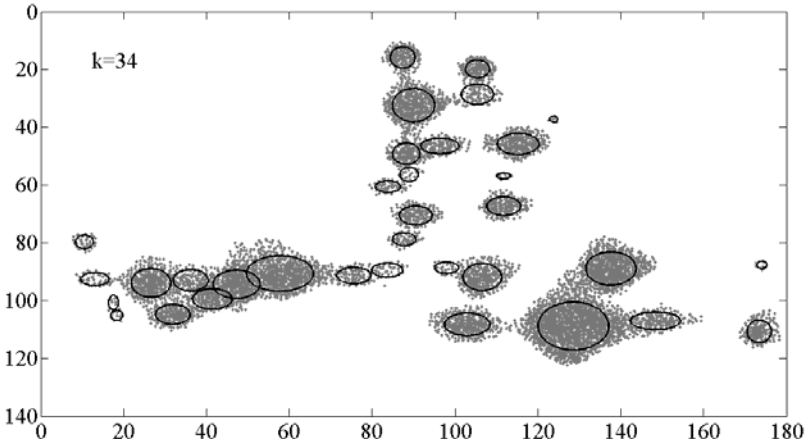
Spot centers estimation
(35 centers)



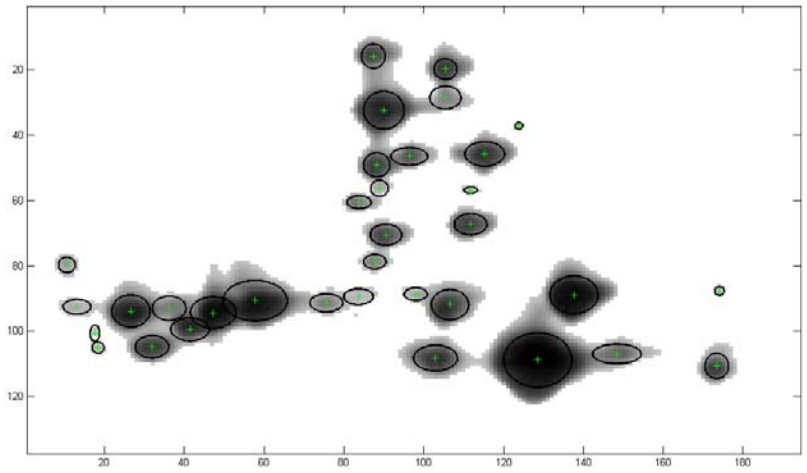
Spot modeling initialization
(35 components)



Spot modeling result
(34 models)



Spots detected (34)



Spot detection evaluation - Comparison to PDQuest

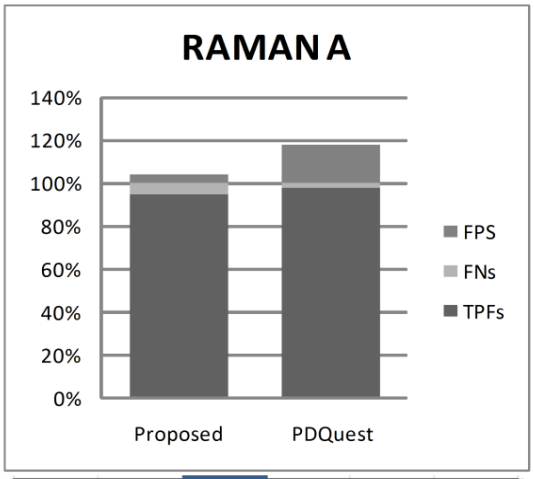
Four real 2DGE images with ground truth known

1. High TP Fraction > 91%
2. Higher precision than PDQuest
   Less false positive spots

➢F-measure:

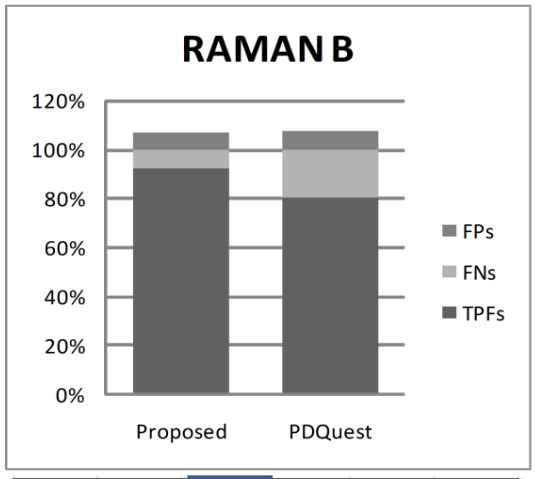
$$F = 2 * \frac{PPV * TPF}{PPV + TPF}$$

	F-Measure (PDQuest)	F-Measure (developed method)
RamanA	90.46%	95.37%
RamanB	85.85%	92.35%
DP03031	81.54%	86.56%
DP03041	81.60%	89.19%



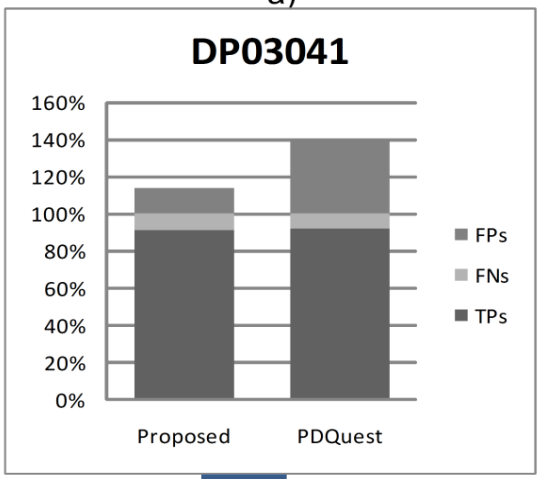
GelA	TPF	PPVs	TPs	FNs	FPs
Proposed	94.83%	95.96%	927	51	39
PDQuest	97.57%	84.31%	962	24	179

a)



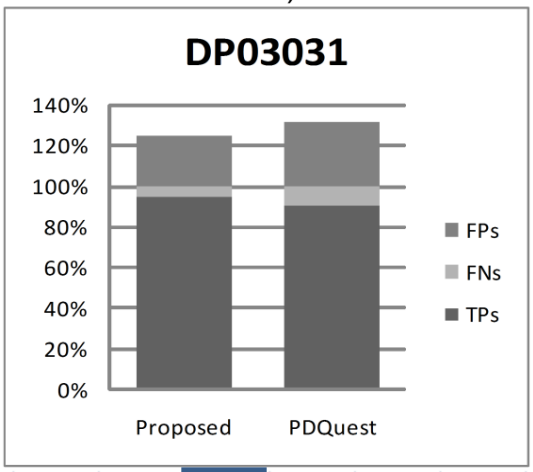
GelB	TPF	PPVs	TPs	FNs	FPs
Proposed	92.13%	93.34%	1135	97	81
PDQuest	80.76%	91.62%	995	237	91

b)



DP03041	TPF	PPVs	TPs	FNs	FPs
Proposed	91.44%	87.05%	598	56	89
PDQuest	92.05%	69.43%	602	52	265

c)



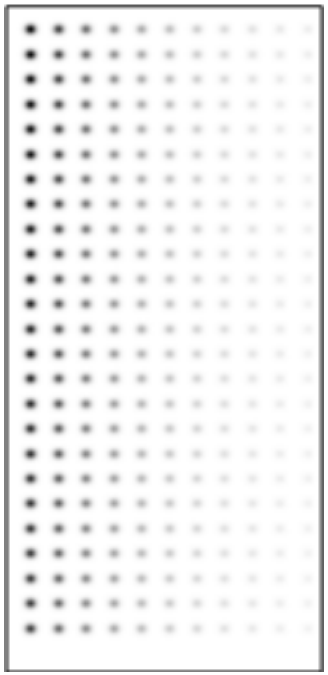
DP03031	TPF	PPVs	TPs	FNs	FPs
Proposed	94.81%	79.63%	512	28	131
PDQuest	90.37%	74.28%	488	52	169

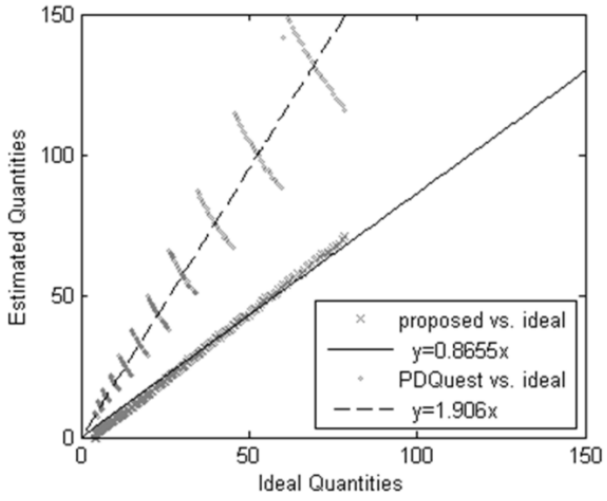
d)

Spot quantitation evaluation

Using one synthetic image, noise σ_n = *10, 20 και 30.*

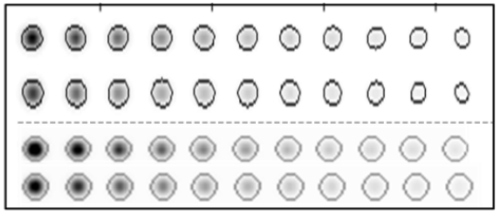
- Estimated spot volumes are closer to the real ones (*a* closer to unity).
- The standard deviation and RMSE are much smaller than PDQuest



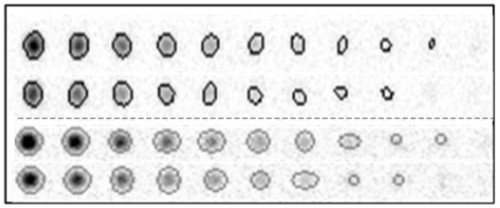


a)

c)



d)



e)

Missed spots (out of 266)	no noise	σ = 10	σ = 20
Proposed	0	50	62
PDQuest	0	101	110

b)

Error Statistics	Error mean (σ=0)	Error STD (σ=0)	RMSE (σ=0)	RMSE (σ=10)	RMSE (σ=20)
Proposed	-4.5793	0.4201	4.7673	27.2257	26.8378
PDQuest	23.6017	6.3825	31.0302	34.661	42.6788

Summary

- The developed contourlets-based image denoising method not only removes noise without distoring the image, but also has a very positive effect on downstream processes
  - i.e. Improves considerably the performance of spot detection, as shown when using the popular PDQuest software package

- The extraction of regions of interest (ROIs) after denoising is shown to be reliable leading to very good results, even in image areas with faint spots or complex spot regions with many overlapping and/or saturated spots.

- The developed ROI extraction method can be used as a mechanism
  - For accepting or rejecting ambiguous spots detected by using a commercial software package,
  - thus accelerating significantly the laborious step of manual spot editing .

Summary

- The end-to-end methodology that was developed for the detection and quantitation of protein spots has been validated with a lot of synthetic and real 2DGE images and it is shown that,
  - not only it improves the precision of spot detection in comparison to commercial software packages,
  - but also the estimates of the spot volumes it provides are more reliable.

- Finally we should note that the whole image analysis pipeline for processing 2DGE images
  - does not require the recalibration of any parameter, every time a new image is presented for processing,
  - and can therefore be fully automated, and support high throughput proteomics image analysis for biomarkers discovery.

Further research directions

- Denoising:
  Using the undecimated Contourlet transform will reduce further distortions due to denoising

- Extracting regions of interest:
  May be improved by using several active contours simultaneously targeting different areas of the image with different characteristics in terms of spots present.

- Apply the method in other bioimage analysis application domains:
  Cell counting and tracking
  Other ?

.

Questions?

Thank you for your attention!

Any Questions?

This work is part of Research Project 03ED306 implemented within the framework of the "Program of Human Research Manpower Reinforcement" (PENED), co-financed by National funds (Greek Ministry of Development-General Secretariat of Research and Technology) and European Union funds.
The authors acknowledge the support of the funding agency. The opinions expressed in this paper are those of the authors and not necessarily of the funding agency.

Related Publications

- Tsakanikas, P., Manolakos, E. S., Improving 2-DE gel image denoising using Contourlets, *Proteomics* 2009, 9, 3877–3888.

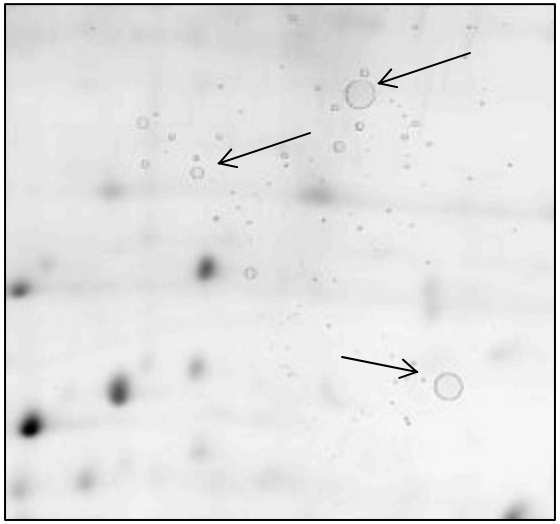
- Tsakanikas, P., Manolakos, E. S., Hierarchical Segmentation of Proteomics gel images using Machine Learning methods, *Proteomics* 2010, submitted, under review.

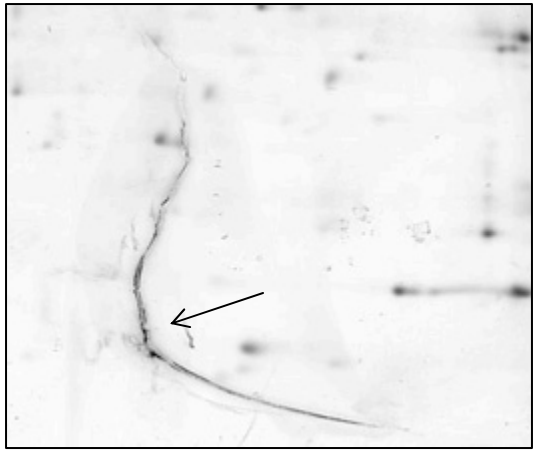
- Tsakanikas, P., Manolakos, E. S., Effective Denoising of 2D Gel Proteomics Images Using Contourlets In Proceedings of the Proceedings *IEEE International Conference on Image Processing* (ICIP) 2007, San Antonio, Texas, USA, September 16-19, 2007, pp. VI: 269-272.

- Tsakanikas, P., Manolakos, E. S., Active Contour Based Segmentation of 2DGE Proteomics Images, In the Proceedings of the *16th European Signal Processing Conference* (EUSIPCO-2008), Lausanne, Switzerland, August 25-29, pp. 83-87, 2008.

Difficulties with 2DGE image analysis

- Bubbles → lead to pseudo-spots
- "Broken" gel areas → streaks, pseudo-spots
- Even human fingerprints…





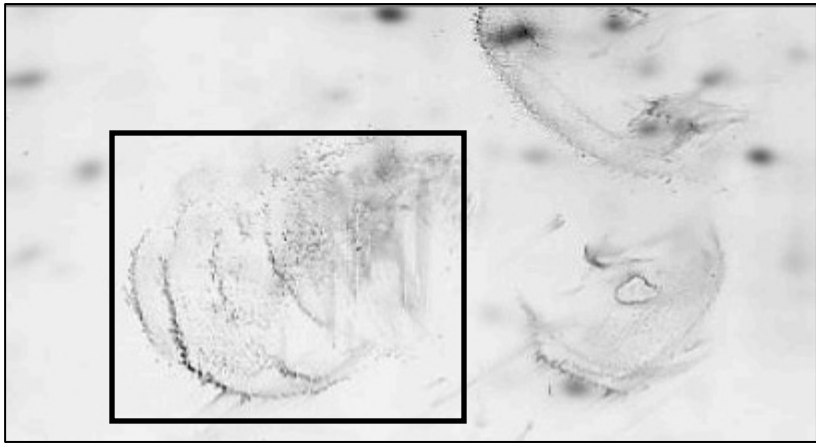


Image denoising

- 2DGE images are typical examples of non-stationary signals
  → extensive variability is observed in spot size, shape and intensity

- The software packages use spatial filters (mean, median, Gaussian etc) to combat noise.

✓However spatial filtering methods distort spot boundaries and in addition alter pixel intensities.

- Recently, wavelet-based methods have been shown to be more effective than spatial filters in denoising 2DGE images [K. Kaczmarek, et. al.]
  However the Wavelet transform:
  - ✓ may represent edges geometry only along 3 main directions (horizontal, vertical, diagonal) and does well only with smooth edges.

  ✓It is a good transform for exploiting point discontinuities (edges) but not of the regularity and smoothness of the spot boundaries.

K. Kaczmarek, et. al. "Preprocessing of two-dimensional gel electrophoresis images", *Proteomics* 2004, *4,* 2377–2389

Contourlets-based denoising

The Contourlet Transform

✓A flexible multirate image transform that can represent spot edges by using contour segments

✓It offers a high degree of directionality and anisotropy

✓Can represent spot boundaries with less coefficients than the Wavelet transform.

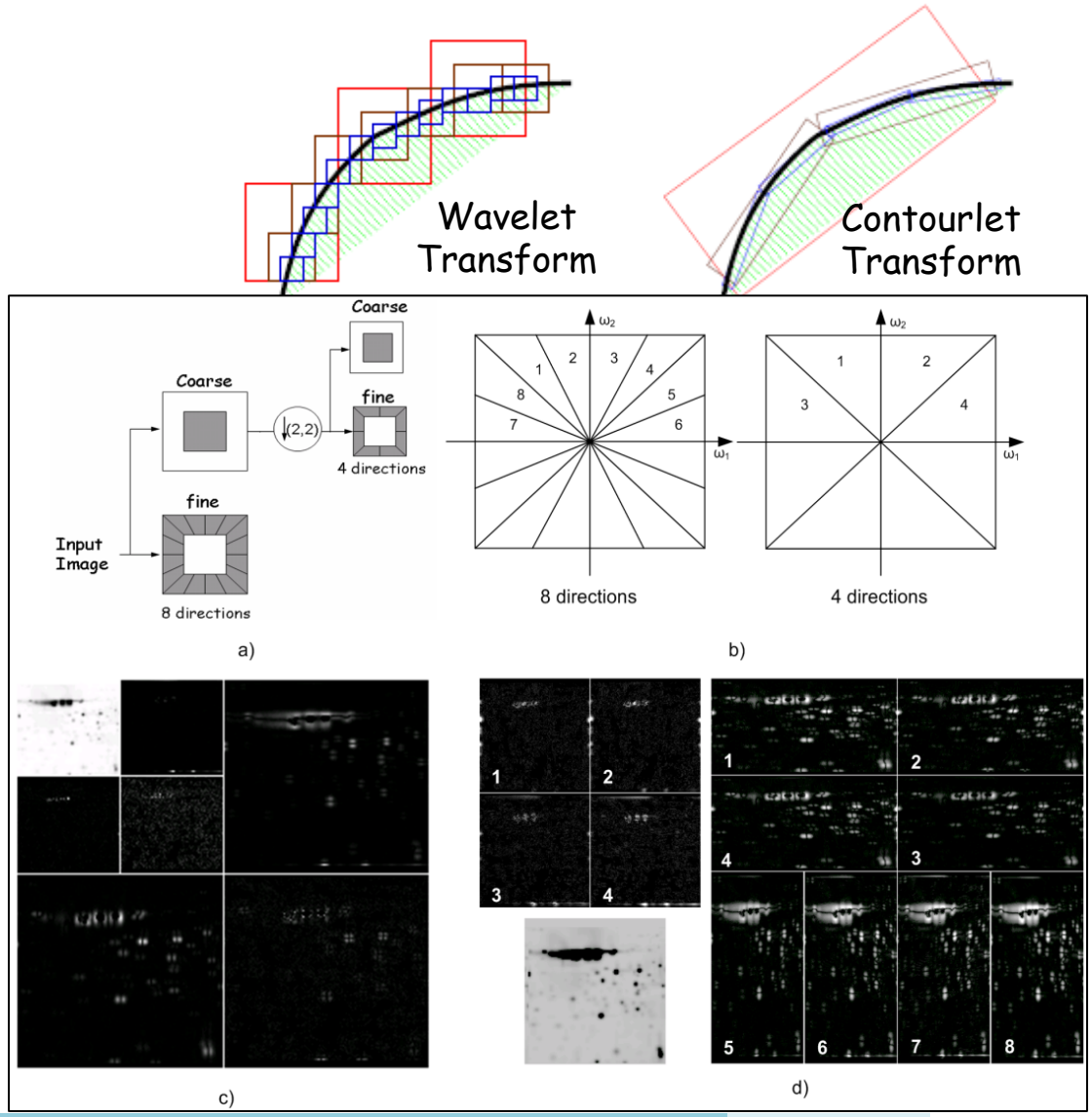


Image denoising

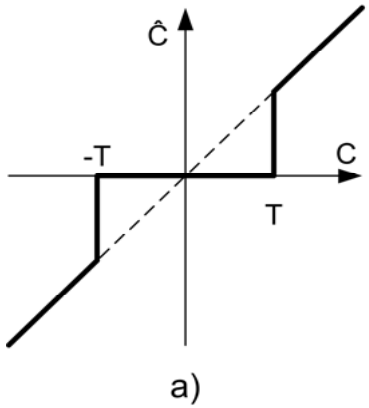
- Multi-resolution image decomposition
- Transform coefficients thresholding
  - ✓Hard thresholding (a)
  - ✓Soft thresholding (b)
  - 1.Bayes threshold:

$$T_s = \frac{\sigma_n^2}{\sigma_s}$$

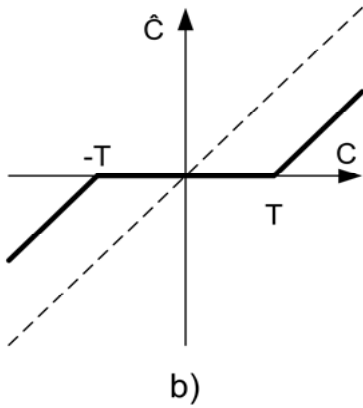
  - 2.Bivariate threshold :

$$\hat{C}_1 = \frac{\left(\sqrt{C_1^2 + C_2^2} - \sqrt{3}\frac{\sigma_n^2}{\sigma_s}\right)_+}{\sqrt{C_1^2 + C_2^2}} \cdot C_1$$

- Image reconstruction



a)



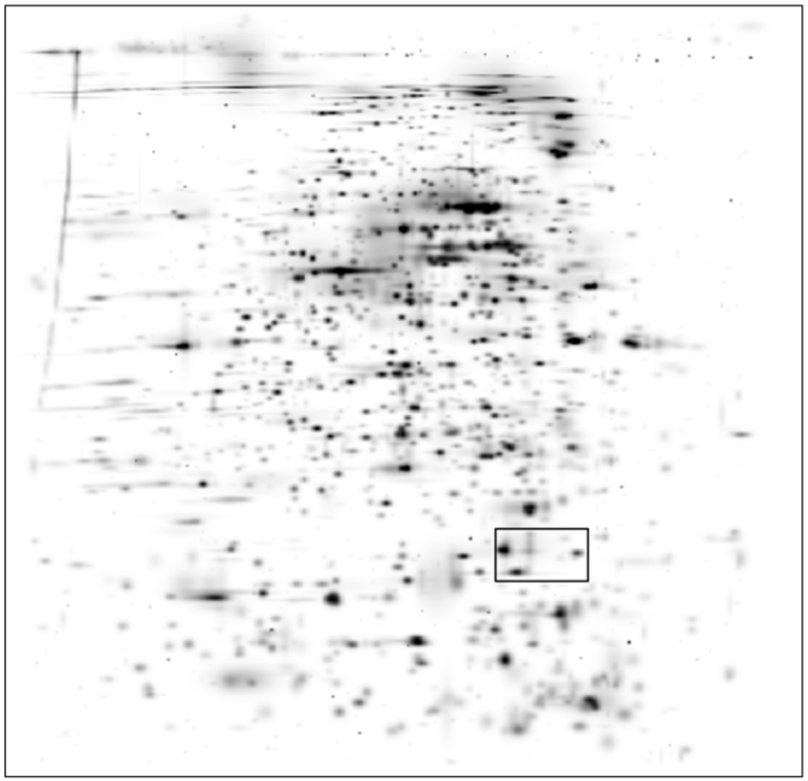
b)

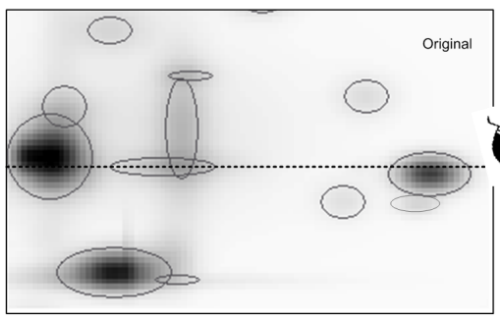
- 4 transform-thresholding combinations have been evaluated
  1. WT-Bayes,
  2. CT-Bayes,
  3. WT-Bivariate,
  4. CT-Bivariate.

Already found to be better than spatial filtering*

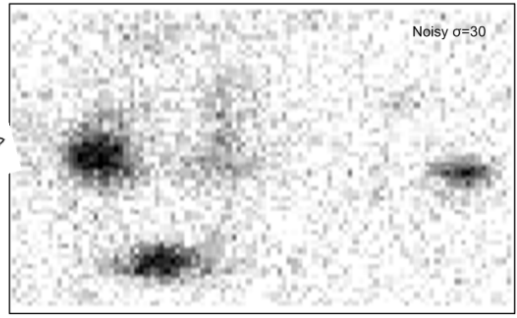
K. Kaczmarek, et. al. “Preprocessing of two-dimensional gel electrophoresis images”, *Proteomics* 2004, *4,* 2377–2389

Effect of denoising on spot detection

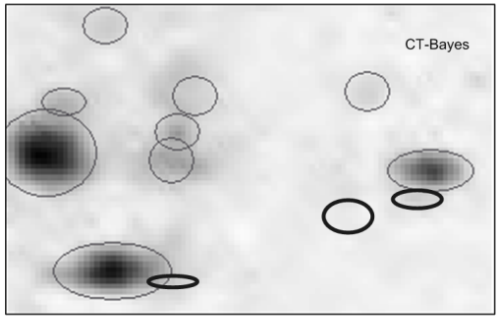

a)



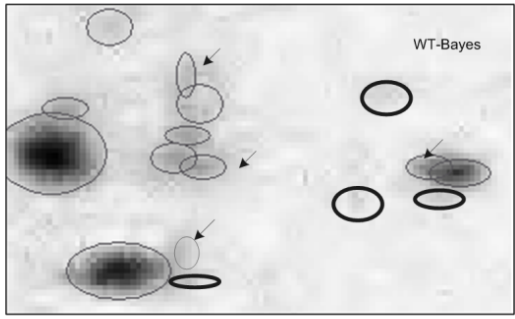
b)



c)



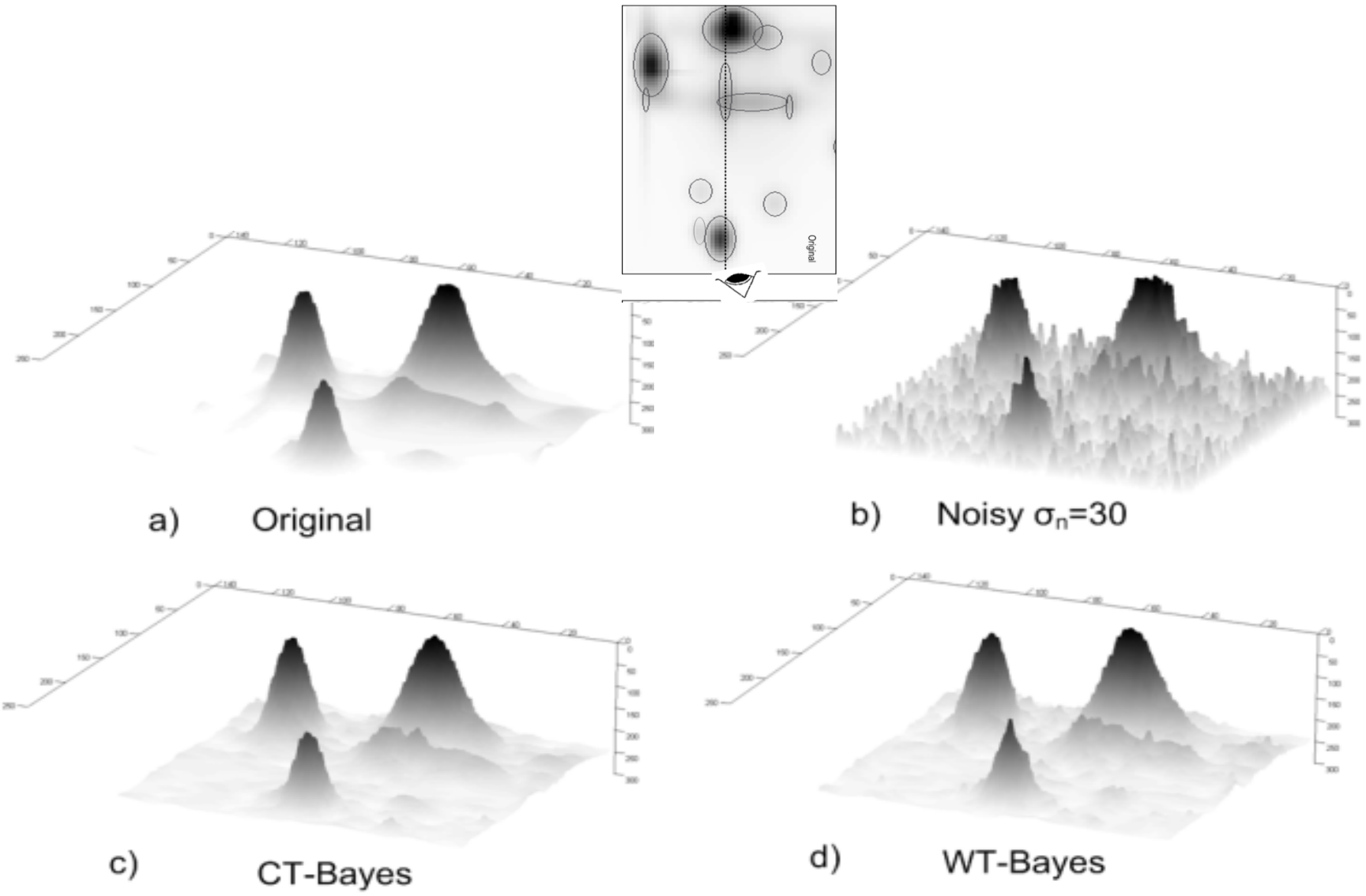
d)



e)

- ✓Small arrows → Introduced extraneous spots (FP)
- ✓ Dark ellipses → Missed spots (FN)

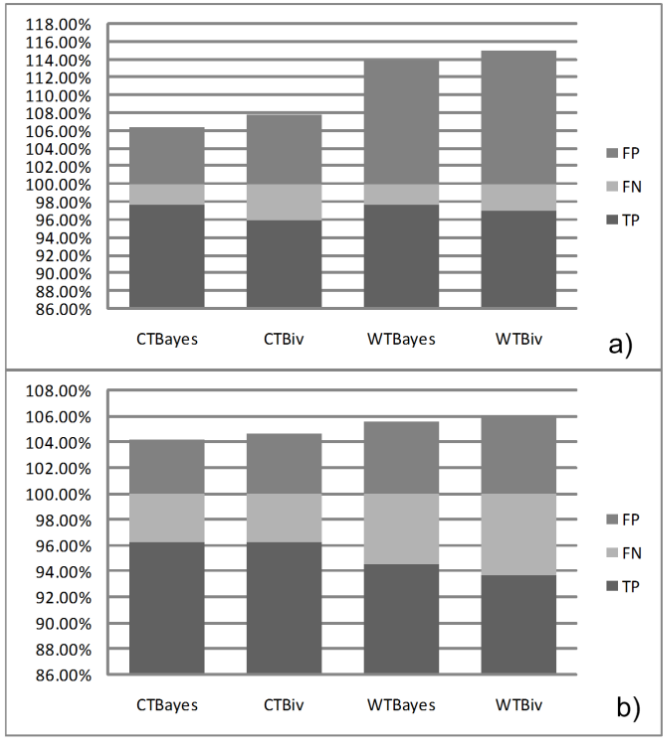
Denoising effect on spot detection (cont..)



Denoising assessment – Effect on spot detection

2 real images with ground truth known (Raman's dataset)

- Reduced percentage of false positive spots, 4-8% instead of 6-15% with the best wavelet based denoising method
- Reduced number of missed "faint" spots (FNs)
- PDQuest used for spot detection in all cases



$$TPF = \frac{True\ Positives}{True}$$

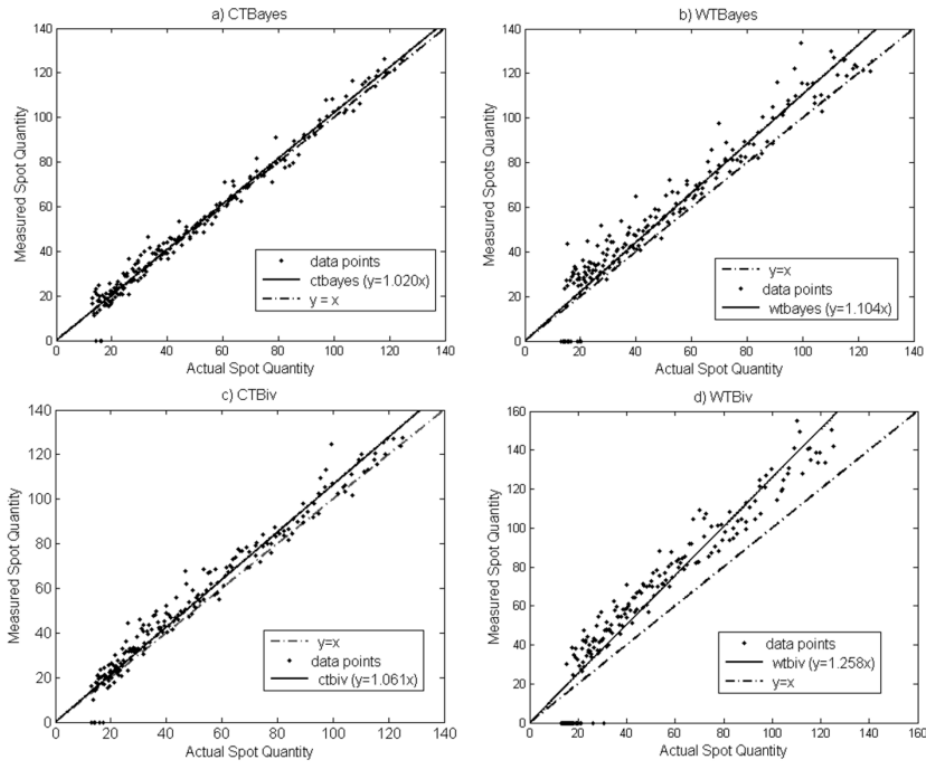
GelA	TPF	FNs	FPs
CTBayes	97.77%	22	63
CTBiv	96.04%	39	77
WTBayes	97.67%	23	137
WTBiv	97.16%	28	148
GelB	TPF	FNs	FPs
CTBayes	96.33%	51	58
CTBiv	96.33%	51	64
WTBayes	94.53%	76	78
WTBiv	93.74%	87	84

c)

Denoising – effect on spot volume estimation

3 synthetic images (Roger's dataset), effect on spot volumes estimation

Noise (σ_n)	10		20		30	
	% error	Standard deviation	% error	Standard deviation	% error	Standard deviation
Quant1	Fixed spot size but fainting maximal intensity					
CTBayes	5.17%	0.0664	8.85%	0.1278	14.25%	0.1946
WTBayes	6.80%	0.1124	23.03%	0.245	32.67%	0.4161
CTBiv	7.48%	0.0855	12.87%	0.1339	25.14%	0.3614
WTBiv	17.79%	0.299	40.14%	0.2196	53.88%	0.3327
Quant2	Image with repeated spot pairs of fixed maximal intensity value but variable degree of spot overlapping					
CTBayes	2.41%	0.0326	3.35%	0.046	5.56%	0.0416
WTBayes	3.17%	0.0396	9.02%	0.0783	11.67%	0.0401
CTBiv	2.42%	0.031	4.87%	0.0534	8.16%	0.0479
WTBiv	5.72%	0.087	17.84%	0.0897	22.42%	0.0407
Quant3	Image with spot pairs of variable spot size and degree of overlapping					
CTBayes	2.36%	0.0404	3.02%	0.0456	5.37%	0.0514
WTBayes	4.89%	0.0479	10.19%	0.0966	14.45%	0.087
CTBiv	3.49%	0.0448	5.53%	0.0563	8.91%	0.0895
WTBiv	7.66%	0.0509	19.39%	0.1213	26.48%	0.1231



	CTBayes	WTBayes	CTBiv	WTBiv
a	1.02	1.104	1.061	1.258
RMSE	3.6	8.894	5.235	13.24
Missed spots	3	17	6	34

Smaller %error and error variability, at all noise levels, and degrees of spot overlapping.

Extracting ROIs – Active Contours

➢ **Active Contrours without edges [T.F. Chan & L.A. Vese]:**

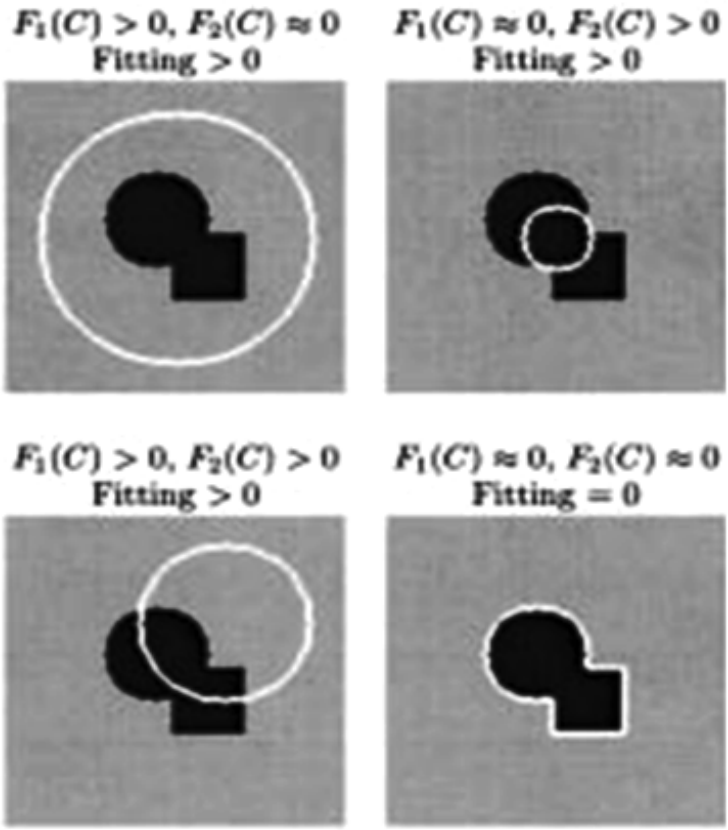
✓ Evolving curve:

$$F_1(C) + F_2(C) = \int_{inside(C)} |u_0(x,y) - c_1| dxdy + \int_{outside(C)} |u_0(x,y) - c_2| dxdy$$

✓ Energy to be minimized

$$E(C, c_1, c_2) = \mu \cdot Length(C) + \nu \cdot Area(C) + \lambda_1 F_1(C) + \lambda_2 F_2(C)$$

✓ Automatic updating of the curve's topology



ROIs extraction – some details

➢ **AC contour Initiallization**

✓ The initial active contours curve is computed automatically by contourlet-based image analysis, by setting to zero the low resolution coefficients. It is possible in this way to find a hyper-surface that includes all spots and is a good starting point for the energy minimization problem.

➢ **Appropriate parameters (fixed for all images):**

- **μ**, controls the importance of the length of the evolving curve in the energy function minimization (μ = 0.006 * 255^2)
- **ν**, controls the importance of the area of the evolving curve in the energy function minimization (ν = 0)
- λ_1, λ_2, their ratio controls which side of the curve has greater importance in the energy function minimization (λ_1=10, λ_2=1)

➢ **Image enhancement:**

✓ The Contourlet transform is also used to enhance the low intensity spots (contrast enhancement) before applying image segmentation.

Denoising evaluation – SNR improvement

Two Datasets used:

✓100 synthetic images (Gaussian spots)

✓8 synthetic images closely resembling real images

[Rogers et al., Proteomics 2003]

Added noise with $\sigma_n = 10, 20, 30$

$$\mathrm{SNR} = 10 \cdot \log_{10} \frac{1}{\mathrm{MSE_N}}$$

$$\mathrm{MSE_N} = \frac{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2}{\sum_{i=1}^{n} (x_i)^2}$$

✓In all cases Contourlet-based denoising performs better than wavelet-based denoising in terms of SNR.

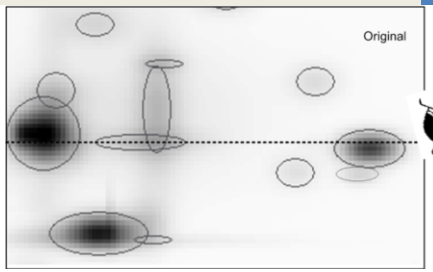
σ_n	WT-Bayes	CT-Bayes	WT-Biv	CT-Biv
10	40.07	40.56	39.58	**40.82**
20	35.59	**36.11**	34.89	36.09
30	33.04	**33.63**	32.30	33.36

Average PSNR, 100 synthetic images Gaussian spots

σ_n	WT-Bayes	CT-Bayes	WT-Biv	CT-Biv
10	41.38	41.77	40.39	**42.33**
20	37.32	37.85	36.11	**38.27**
30	35.00	35.42	33.70	**35.99**

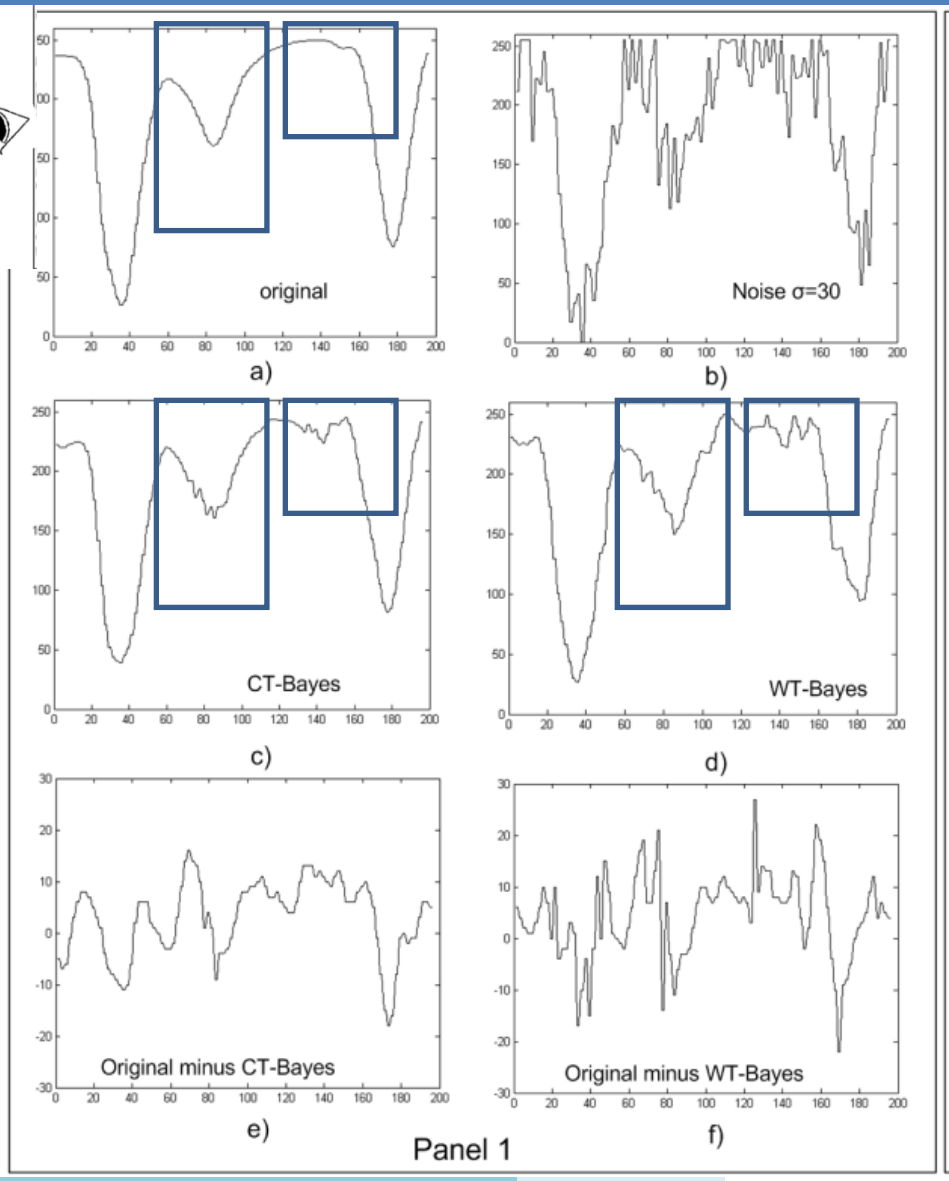
Average PSNR, 8 pseudo-realistic images

Αξιολόγηση αποθορυβοποίησης



- Η αποθορυβοποίηση με βάση τον Contourlet, επιδεικνύει ένα πιο ομαλό προφίλ που προσεγγίζει καλύτερα αυτό της εικόνας πριν την εισαγωγή του θορύβου.

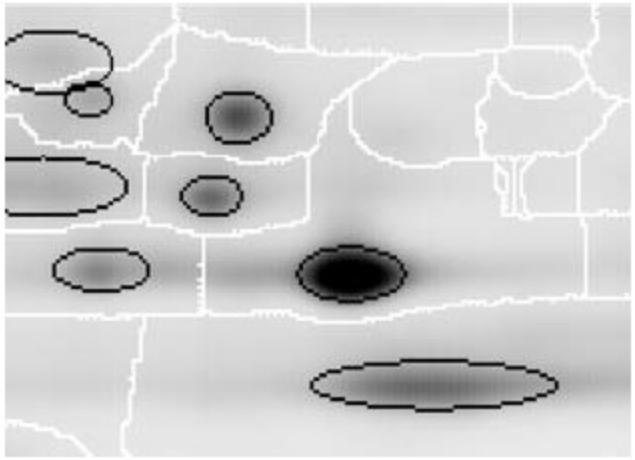
- Η αποθορυβοποίηση με βάση τον Contourlet, εισάγει λιγότερες παραμορφώσεις, ιδιαίτερα τα όρια των κηλίδων.



Κατάτμηση Εικόνας και εξαγωγή περιοχών ενδιαφέροντος

- Η κατάτμηση των εικόνων πηκτωμάτων προυποθέτει τον διαχωρισμό του φόντου από τις περιοχές που εμπεριέχουν πρωτεϊνικές κηλίδες (περιοχές ενδιαφέροντος)

- Προηγούμενες Μέθοδοι:
  - ✓ Βαθμιδωτά ρυθμιζόμενη κατωφλιωποίηση (stepwise)
    - ιδιαίτερα ευαίσθητη στον θόρυβο και στα «ψευδή» αντικείμενα,
    - χρειάζεται επιπλέον κριτήρια για την αποδοχή ή απόρριψη των τελικών περιοχών
  - ✓ Μέθοδος δεύτερης παραγώγους
    - ιδιαίτερα ευαίσθητη στον θόρυβο και στα «ψευδή» αντικείμενα
    - Προκαλεί συρρίκνωση των ορίων των κηλίδων
  - ✓ Μετασχηματισμός Watershed
    - παρουσιάζει υπερβολική κατάτμηση της εικόνας
    - δεν είναι τετριμένη η διαδικασία καθορισμού σημείων ελέγχου
  - ✓ Μοντελοποίηση με στατιστικά μοντέλα
    - Παρουσιάζουν δυσκολία στην εφαρμογή τους καθώς δεν υπάρχει προγενέστερη γνώση του σχήματοσ και του μεγέθους των κηλίδων



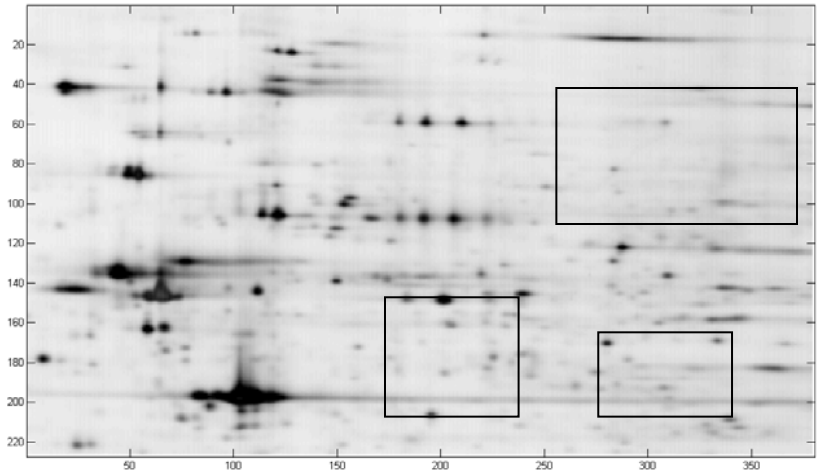
Κατατμημένη εικόνα με στατιστικά μοντέλα (ελλείψεις) και με Watershed (άσπρες γραμμές)

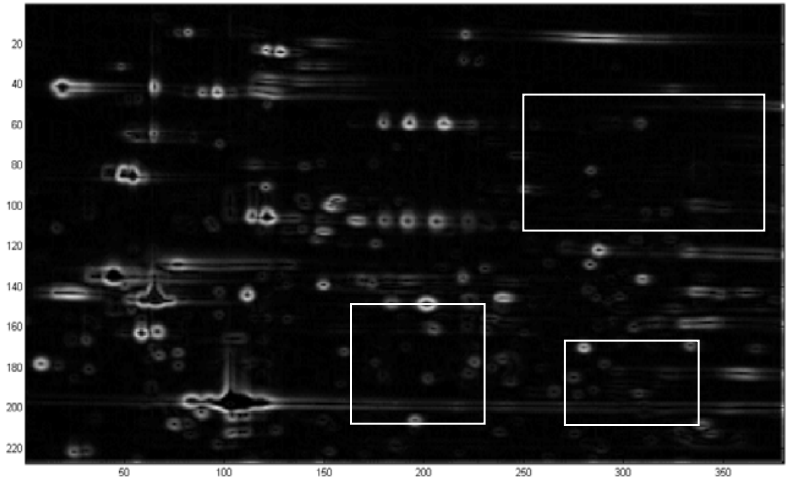
✓Level Set formulation of Active Contours:

✓Automatical topological changes of the curve, which is modeled as a specific level set function of time in a higher dimensional space.

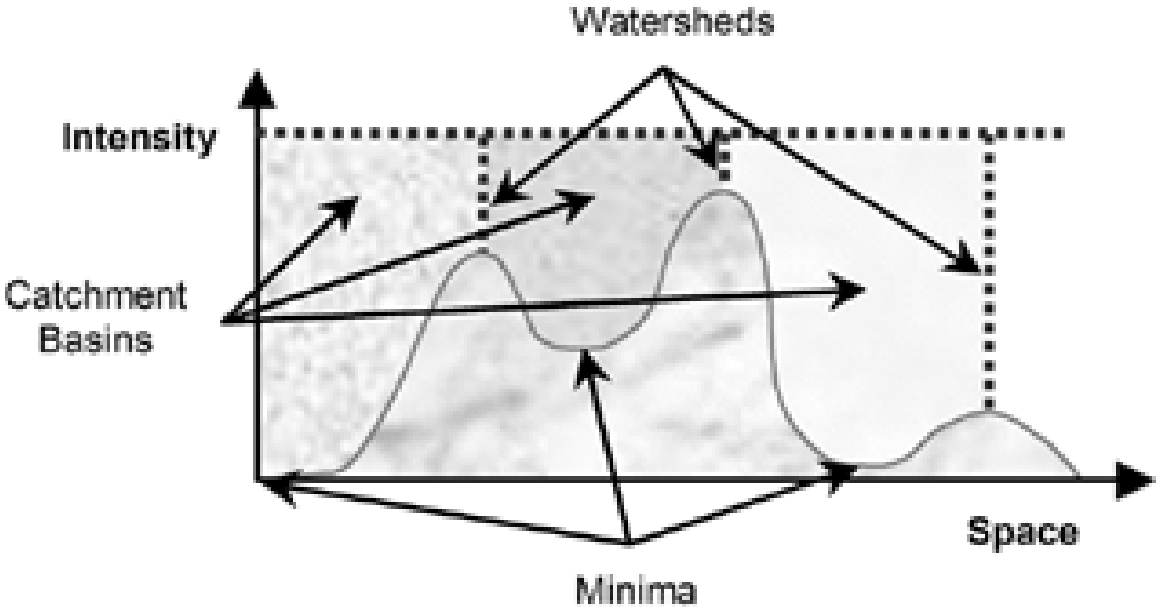
✓ The evolution of the curve relies on an edge function depending on the gradient $|\nabla u_0|$ of the image u_0

- Can detect only objects well defined by their gradient.
- The curves, in practice, is likely to pass inwards the spots borders, especially the faint ones!!!

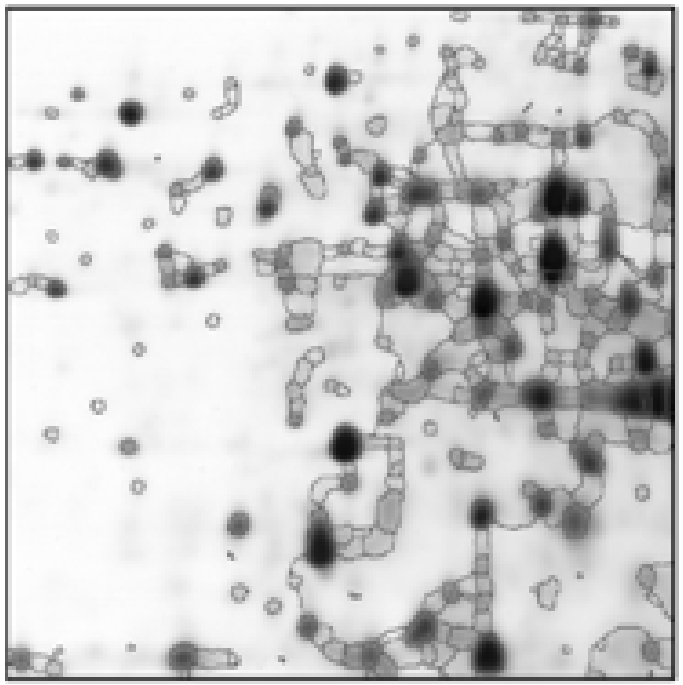




Κατάτμηση Εικόνας και εξαγωγή περιοχών ενδιαφέροντος
Back Slide 1



a)

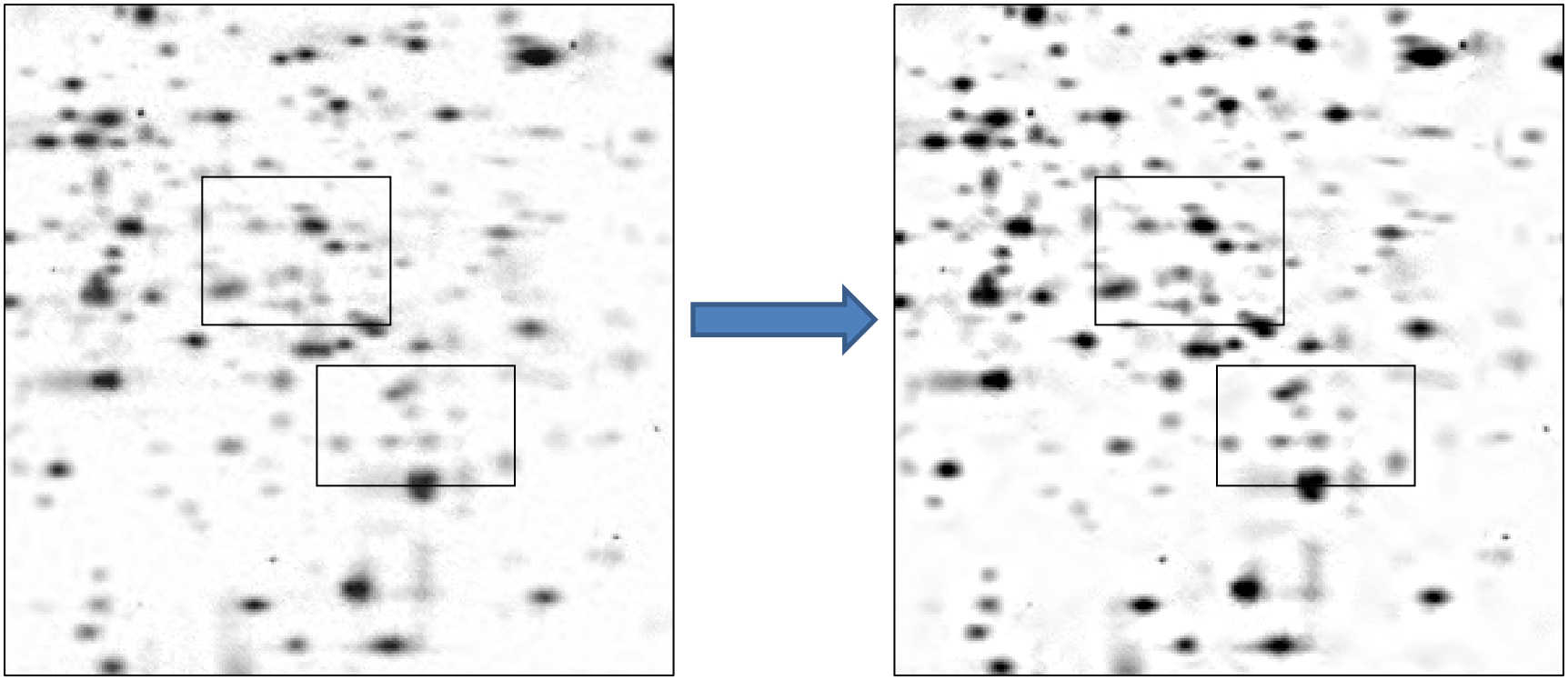


b)

Κατάτμηση Εικόνας και εξαγωγή περιοχών ενδιαφέροντος

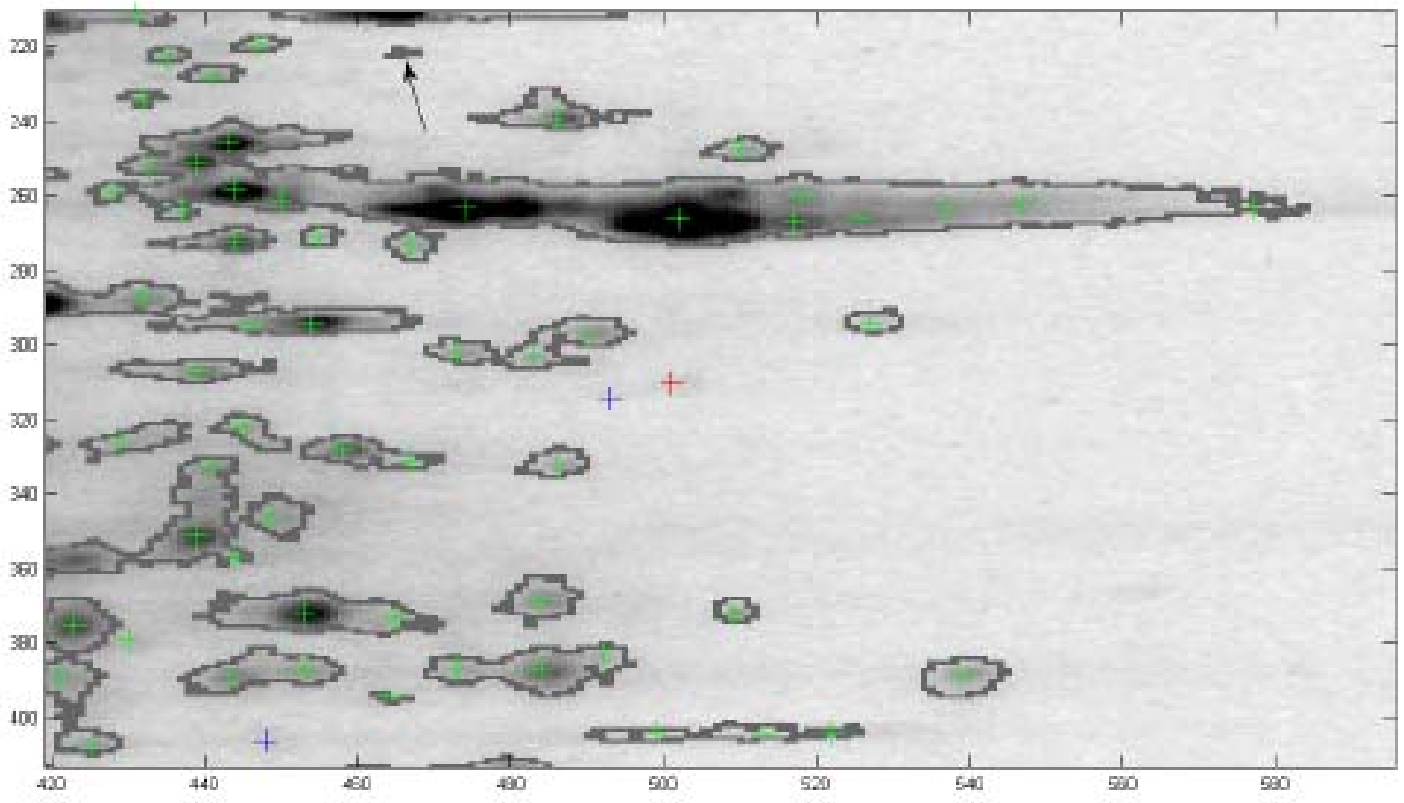
➢Η εφαρμογή της ακόλουθης σχέσης στους συντελεστές (u) των μεσαίων συχνότητας της εικόνας (μετά από τον εφαρμογή του Contourlet) έχει ως αποτέλεσμα την «ενίσχυση» (enhancement) των κηλίδων, αποφεύγοντας την «ενίσχυση» του φόντου (χαμηλές συχνότητες) και του θορύβου (υψηλή συχνότητες):

$$E(u) = u \cdot sign(u) \cdot \tanh(b \cdot n) \cdot (1 + c \cdot \exp(-n^2))$$



E. Nezhadarya, M.B. Shamsollahi, "Image Contrast Enhancement by Contourlet Transform", 2006, pp. 81-84.

Κατάτμηση Εικόνας και εξαγωγή περιοχών ενδιαφέροντος
Back Slide 2

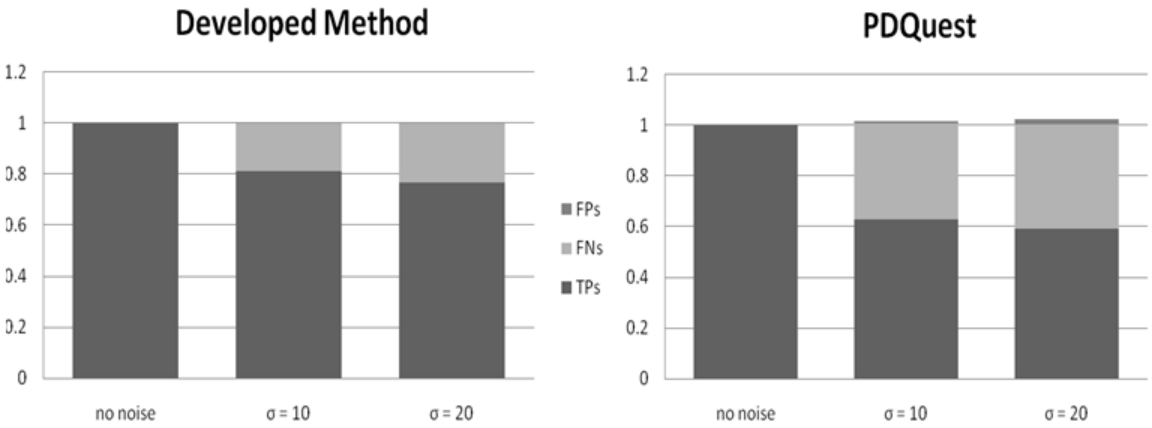


Κατάτμηση Εικόνας και εξαγωγή περιοχών ενδιαφέροντος
Back Slide 3

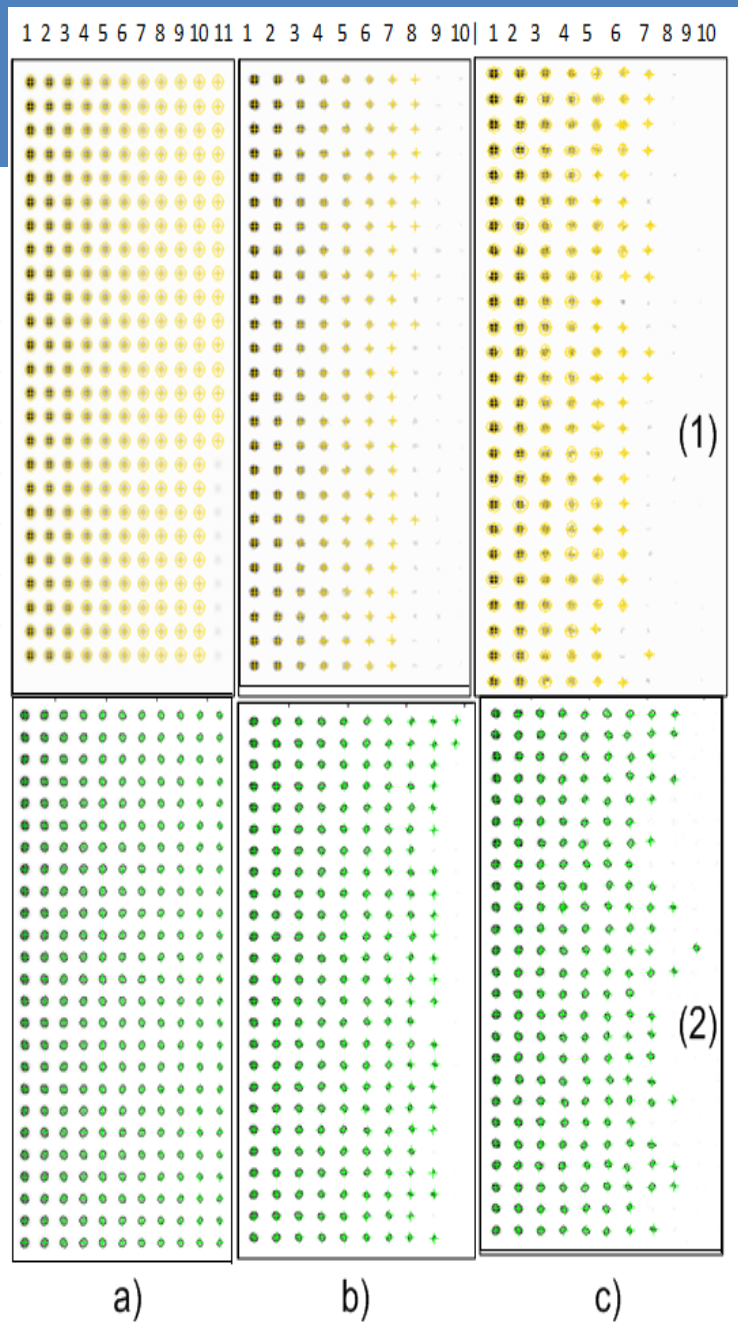
- Ο λόγος των μη ανιχνευθέντων κηλίδων σε σχέση με αυτές του PDQuest είναι ιδιαίτερα μικρός
- Ο λόγος των κοινών ανιχνευθέντων κηλίδων είναι της τάξης του >90%.
- Η ευαισθησία (Sensitivity) είναι πάνω από 95%.
- Η εμπιστοσύνη είναι πάνω από 96% για όλες τις εικόνες.

Image	PDQ	PDQ/AC	% PDQ/AC	PDQ/nAC	% PDQ/nAC	nPDQ/AC	%nPDQ/AC
1a	1112	1090	98,02%	22	1,98%	13	1,19%
2a	1315	1283	97,57%	32	2,43%	7	0,55%
MP1	262	256	97,71%	6	2,29%	13	5,08%
MP2	265	242	91,32%	23	8,68%	11	4,55%
MP3	227	223	98,24%	4	1,76%	24	10,76%
Rj1	146	123	84,25%	23	15,75%	4	3,25%
RGA	948	919	96,94%	29	3,06%	9	0,98%
RGB	1040	1018	97,88%	19	1,83%	40	3,93%

Image	PDQ	AC/PDQ (TP_1)	PDQ/nAC		nPDQ/AC		S	C
			FN	TN	FP	TP_2		
1a	1112	1090	5	17	5	8	99,55%	99,55%
2a	1315	1283	9	23	5	2	99,30%	99,61%
MP1	262	256	2	4	10	3	99,23%	96,28%
MP2	265	242	14	9	4	7	94,68%	98,42%
MP3	227	223	1	3	6	18	99,59%	97,57%
Rj1	146	123	11	12	1	3	91,97%	99,21%
RGA	948	919	20	9	6	3	97,88%	99,35%
RGB	1040	1018	12	7	13	27	98,86%	98,77%



Detection Method	Noise STD	$\sigma = 0$	$\sigma = 10$	$\sigma = 20$
Developed	TPF	1	0.81	0.77
	TPs	266	216	204
	FNs	0	50	62
	FPs	0	0	0
PDQuest	TPF	1	0.63	0.59
	TPs	266	167	157
	FNs	0	101	110
	FPs	0	2	5



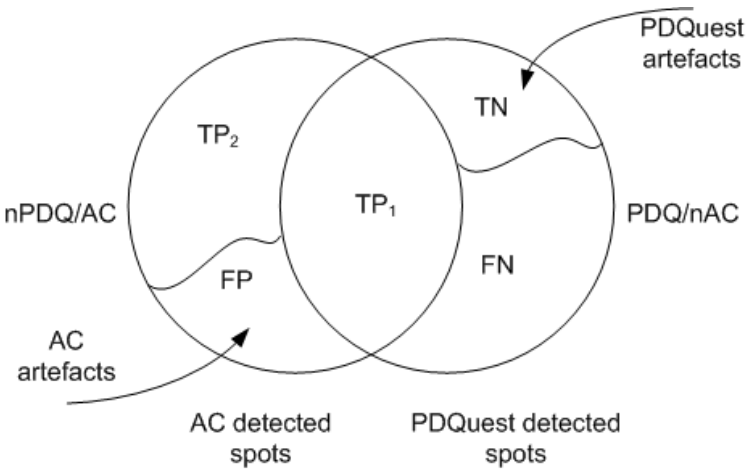
- Υλικά για την αξιολόγηση
  - ✓Έξι συνθετικές εικόνες μεγέθους 1024x1024, 8-bit κλίμακας της γκρι.
  - ✓Δυό πραγματικές εικόνες.
- Μέθοδοι αξιολόγησης
  - ✓ σύγκριση αποτελεσμάτων με αυτά από το PDQuest v7.2.0 in terms of TP, FP, TN and FN.
  - ✓Use of two metrics:

1. Ευαισθησία (Sensitivity), η οποία αποτιμά την αποτελεσματικότητα εύρεσης των πραγματικών κηλίδων:

$$S = \frac{TP_1 + TP_2}{(TP_1 + TP_2) + FN}$$

2. Εμπιστοσύνη (Confidence), η οποία αποτιμά το ποσοστό των πραγματικών κηλίδων προς το συνολικό αριθμό των κηλίδων που ανιχνεύθηκαν:

$$C = \frac{TP_1 + TP_2}{(TP_1 + TP_2) + FP}$$



Spot detection - Finding the best mixtures model that fits the generated dataset

We use mixtures to:

- ✓ To capture the generative mechanism of the data from the pixel intensities
- ✓ To find the number of Gaussian components (underlying spots) that best explains the generated data
- ✓ To extract the parameters of these components (underlying spots) in an unsupervised learning manner.

Algorithm details:

Initialization: 1-NN classification of generated data points

Initial placement of component centers = Estimated spot centers after the hierarchical clustering

Initial mixture coefficients = proportion of points belonging to each estimated center

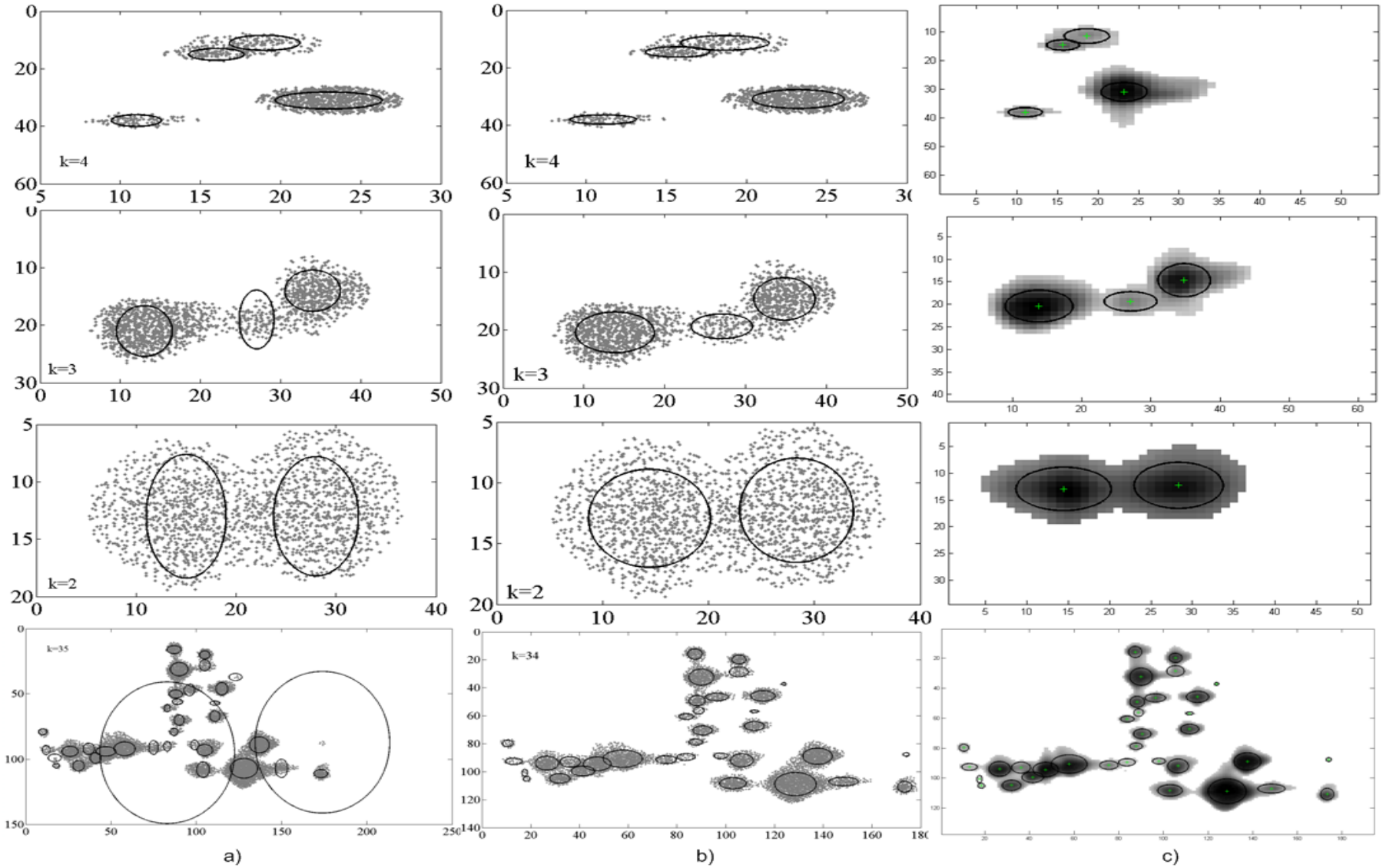
Initial covariance matrices= Sample covariance matrices based on the 1NN classification results

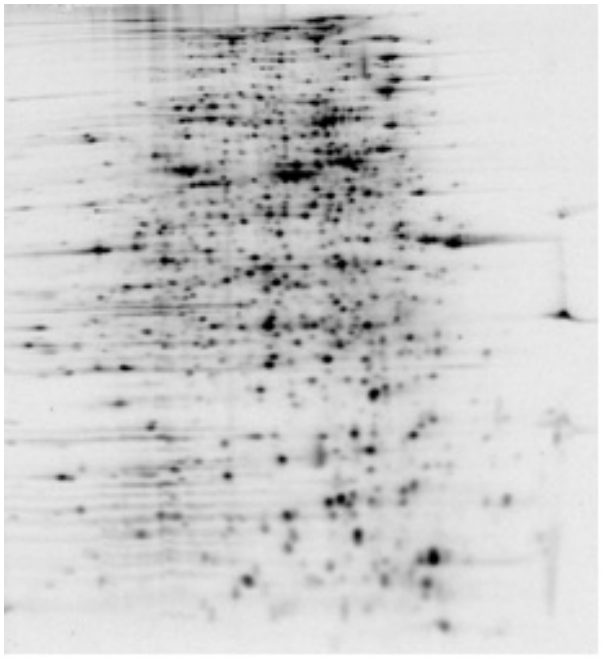
Use the EM algorithm to estimate the mixture model parameters

Use the minimum message length criterion (MML) to identify the number of components in the mixture giving the best results

It is possible to reject spot centers!

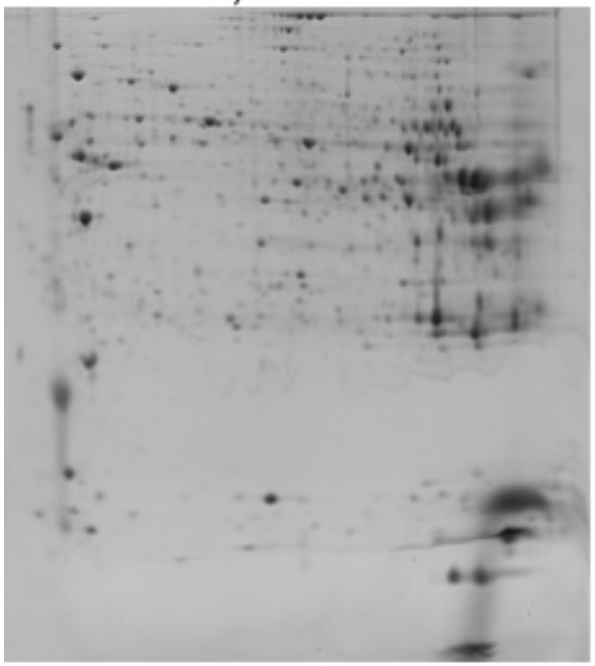
Προσαρμογή μίξης μοντέλων και τελική επιλογή του αριθμού των στοιχείων της μίξης



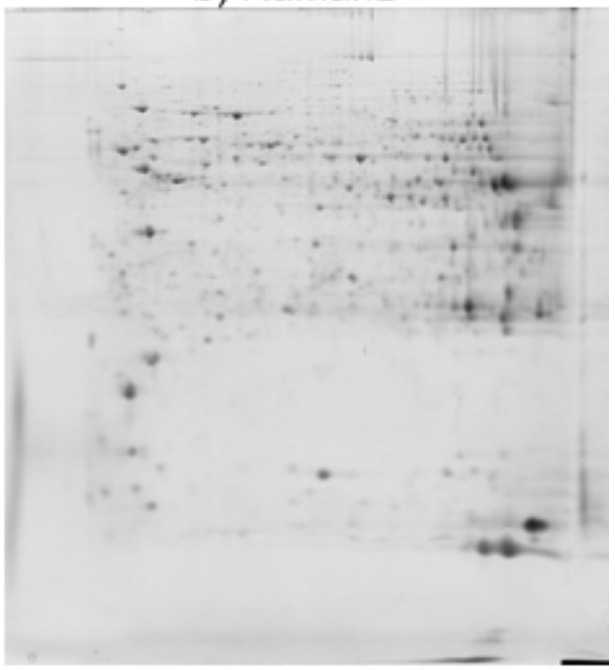


a) RamanA

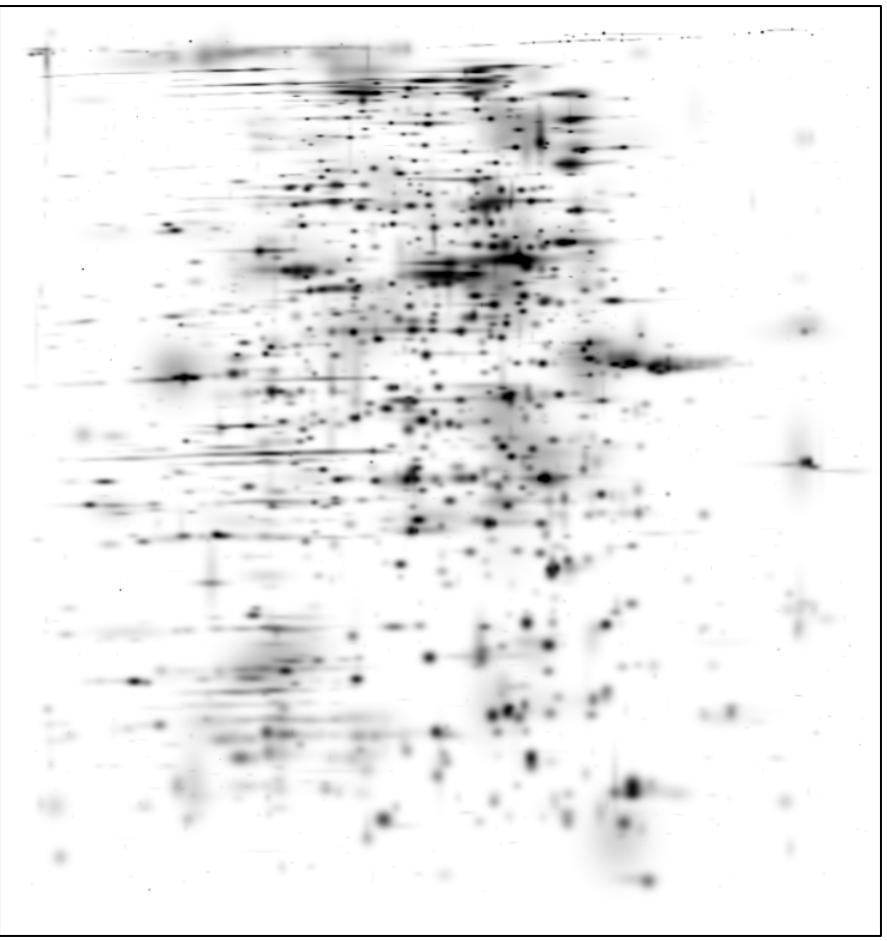
b) RamanB

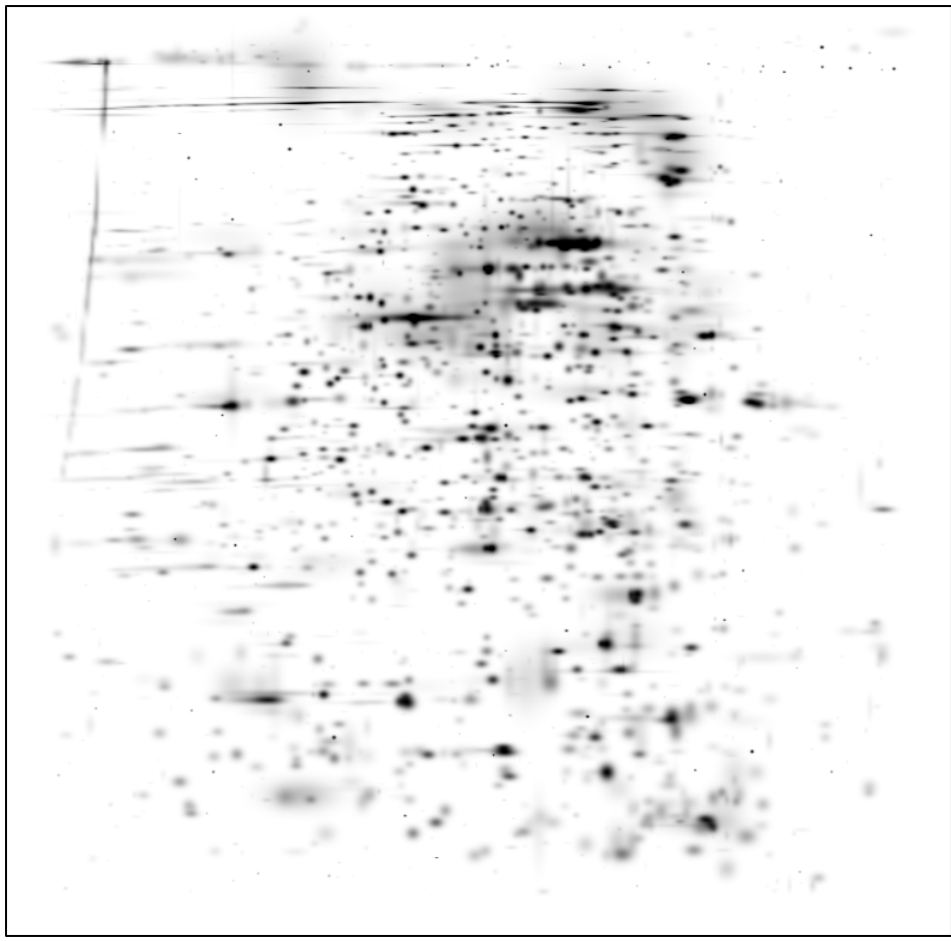


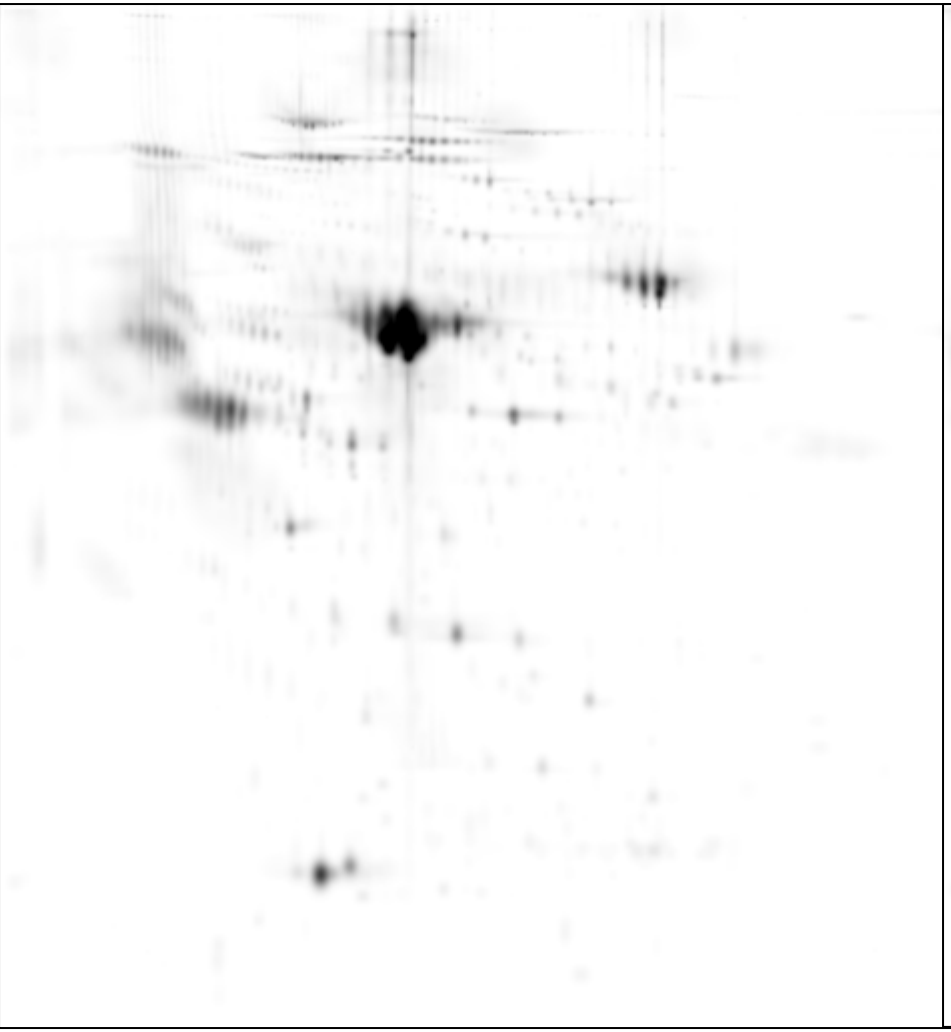
c) DP03041

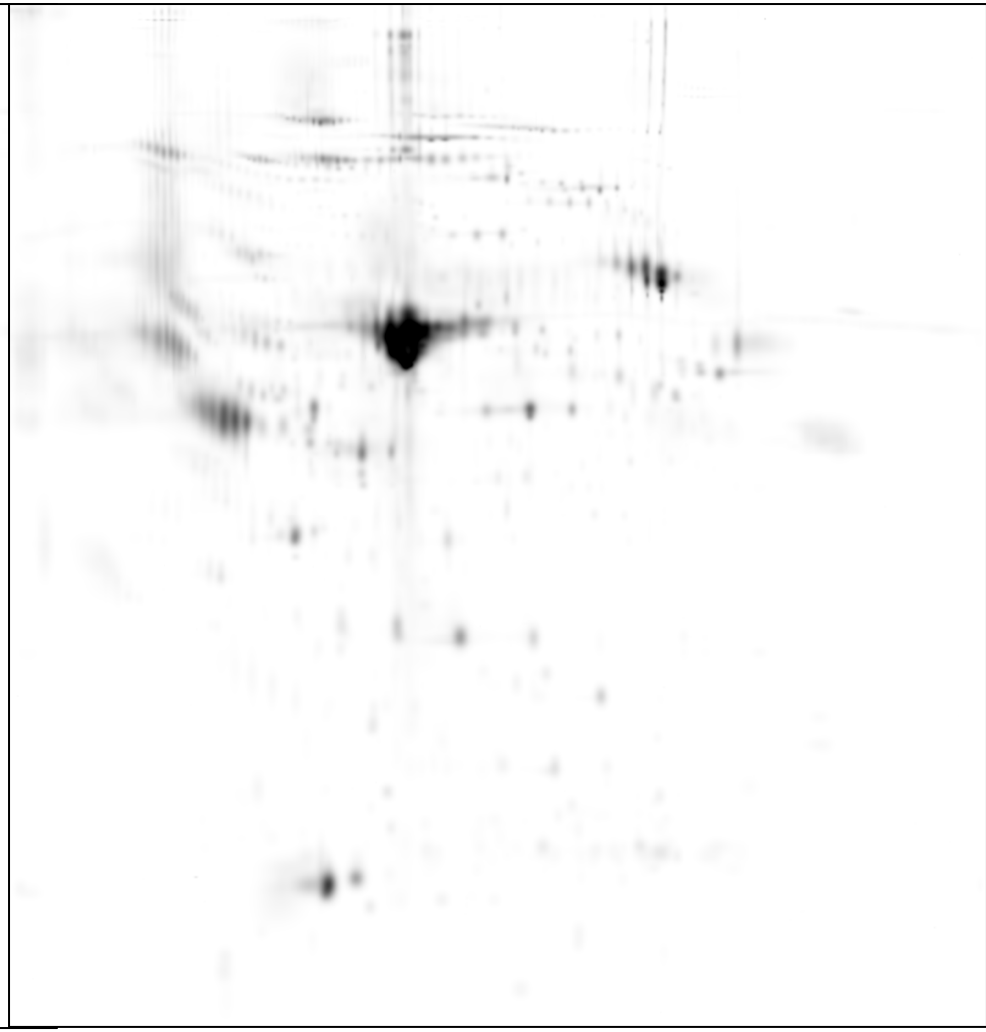


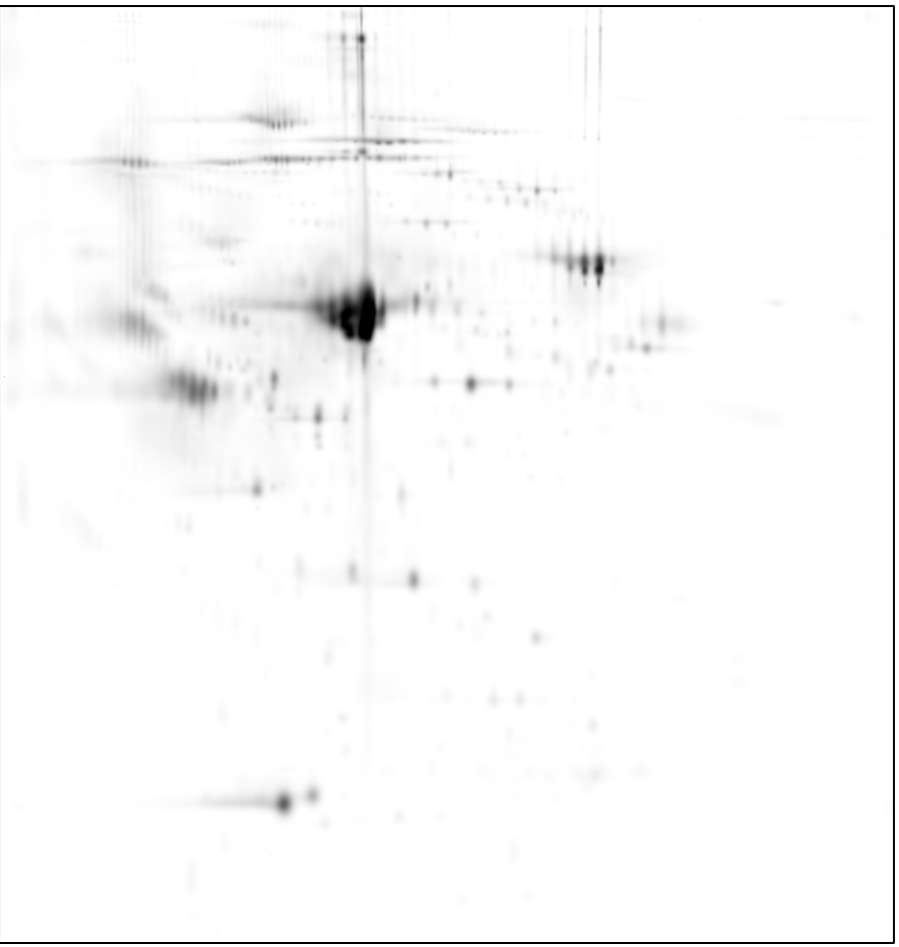
d) DP03031

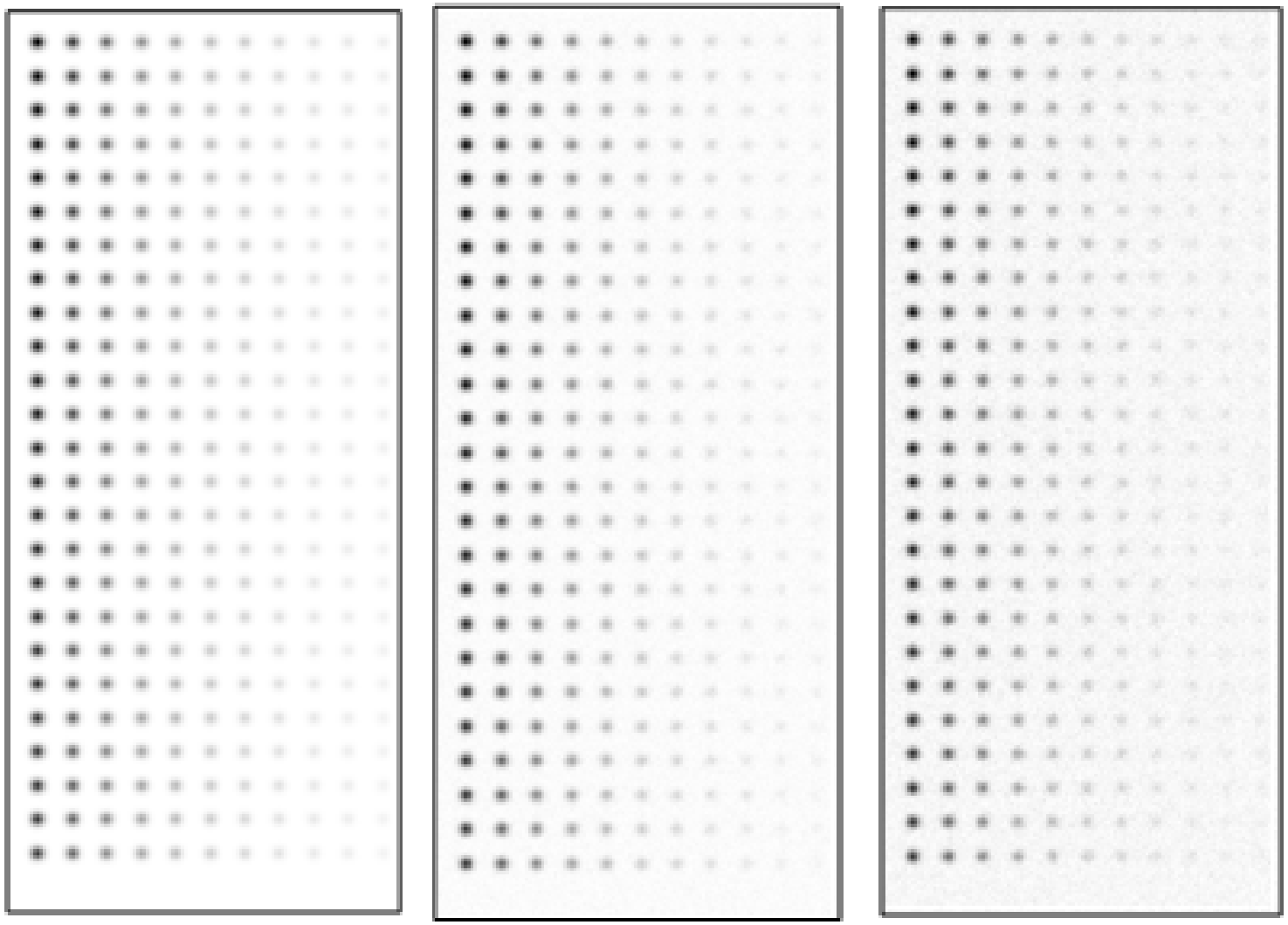


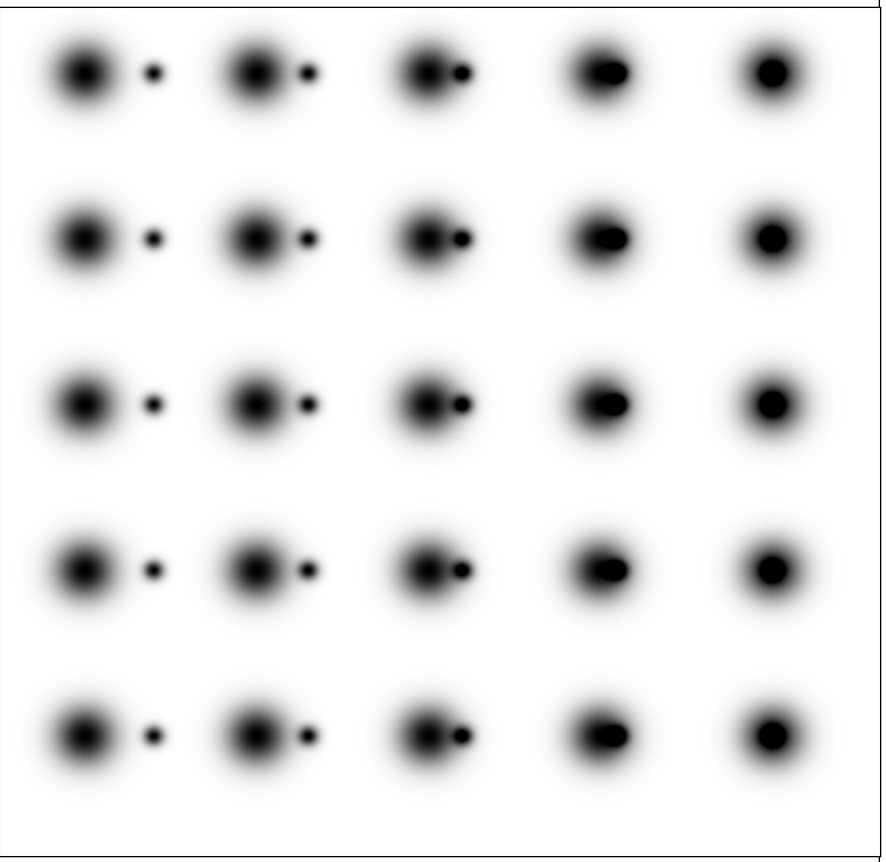


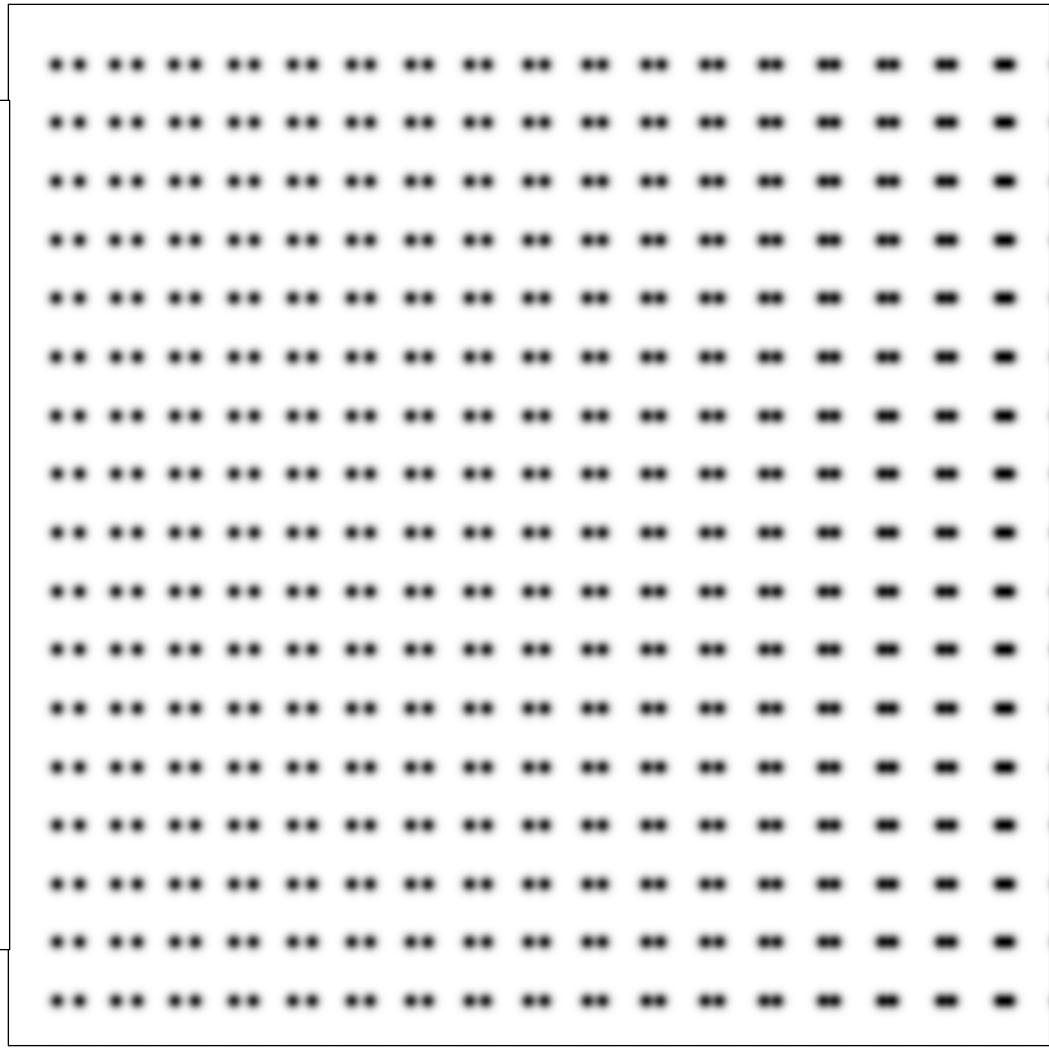







a)
b)
c)





ΑΘΗΝΑ 2010

Δισδιάστατη Ηλεκτροφόρηση

1. Διαχωρισμός των πρωτεϊνών

   - Διάσταση 1: Διαχωρισμός σύμφωνα με το ισοηλεκτρικό σημείο
   - Διάσταση 2: Διαχωρισμός σύμφωνα με το μοριακό βάρος (MW)

2. «Βάψιμο» του πηκτώματος, ώστε να είναι δυνατή η απεικόνιση του σε ψηφιακή εικόνα (2D *gel electrophoresis image*).

3. Εφαρμογή μεθόδων επεξεργασίας εικόνας
   - Ανίχνευση κηλίδων
   - Μέτρηση της έκφρασης των πρωτεϊνών