

# Η ΣΥΜΒΟΛΗ ΤΗΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΔΙΣΔΙΑΣΤΑΤΩΝ ΕΙΚΟΝΩΝ ΗΛΕΚΤΡΟΦΟΡΗΣΗΣ ΣΤΟΝ ΚΛΑΔΟ ΤΗΣ ΠΡΩΤΕΟΜΙΚΗΣ

**Κωστοπούλου Ε.\*, Ζαχαρία Ε., Μαρούλης Δ.**

Τμήμα Πληροφορικής και Τηλεπικοινωνιών, ΕΚΠΑ, Αθήνα

ikostop@di.uoa.gr

## Περίληψη

Η παρούσα εργασία παρουσιάζει έναν αλγόριθμο επεξεργασίας εικόνας για τον αυτόματο εντοπισμό και την κατάτμηση πρωτεϊνικών κηλίδων σε πηκτώματα δισδιάστατης ηλεκτροφόρησης. Η προτεινόμενη μέθοδος χρησιμοποιεί δισδιάστατα ιστογράμματα, αλγόριθμους κατωφλίωσης καθώς και υδροκριτικές γραμμές. Εφαρμόστηκε σε πραγματικές εικόνες πηκτώματος και έδωσε καλύτερα αποτελέσματα σε σύγκριση με σύγχρονα λογισμικά πακέτα και μεθόδους.

## Abstract

This work presents an automatic approach for protein spot detection and segmentation on gel electrophoresis images. The proposed approach utilizes 2D histograms, thresholding techniques as well as watersheds. The experiments conducted on real gels demonstrate that the proposed approach detects and segments effectively the protein spots and outperforms top of the art software packages and techniques.

## Λέξεις Κλειδιά

*Πρωτεομική, πηκτώματα δισδιάστατης ηλεκτροφόρησης, εντοπισμός πρωτεϊνών, κατάτμηση πρωτεϊνών*

## Εισαγωγή

Τα κύτταρα αποτελούν το κύριο δομικό στοιχείο κάθε οργανισμού. Το DNA των κυττάρων είναι φορέας των γενετικών πληροφοριών του οργανισμού. Σύμφωνα με το κεντρικό δόγμα της βιολογίας το DNA μεταγράφεται σε RNA το οποίο στην συνέχεια μεταφράζεται σε πρωτεΐνες. Αλλαγές στα επίπεδα των πρωτεϊνών έχουν συνδεθεί με αρκετές ασθένειες, όπως για παράδειγμα, η παθολογία του καρκίνου η οποία είναι συνδεδεμένη με την υπερέκφραση πρωτεϊνών (Aamodt R., Bondi J., Andersen S.N., Bakka A., Bukholm G. κ.α., 2010). Επιπρόσθετα, συχνά ασθενείς που φαινομενικά πάσχουν από την ίδια ασθένεια και υποβάλλονται στην ίδια φαρμακευτική αγωγή αντιδρούν με εντελώς διαφορετικό τρόπο σε αυτήν. Αυτό συμβαίνει γιατί λόγω του γονότυπου των ασθενών, πολλές ασθένειες μπορεί να έχουν υπότυπους που ποικίλλουν ως προς την κλινική τους πορεία και την πρόγνωση (Ruckhäberle E., Karna T., Hankera L., Schwarz J., Schulz-Knappe P., κ.α. 2010). Η μελέτη της έκφρασης του γονότυπου, δηλαδή των πρωτεϊνών που παράγονται σε κάθε κύτταρο, μπορεί να οδηγήσει σε εξατομικευμένες θεραπείες, μεγιστοποιώντας το κλινικό όφελος και ελαχιστοποιώντας τις παρενέργειες, οι οποίες είναι συχνά πολύ ισχυρές στα φάρμακα ευρέους φάσματος (Roti G., Stegmaier K., 2012).

Η πρωτεομική είναι ένας ραγδαία αναπτυσσόμενος ερευνητικός κλάδος, ο οποίος μελετά το πρωτέωμα, δηλαδή το πλήρες συμπλήρωμα των πρωτεϊνών που εκφράζονται σε ένα κύτταρο, ιστό ή οργανισμό, σε μια δεδομένη χρονική στιγμή (Stannard C., Brown L.R., Godovac-Zimmermann J., 2004). Η δισδιάστατη ηλεκτροφόρηση πηκτωμάτων είναι μια καθιερωμένη μέθοδος της πρωτεομικής με την οποία διαχωρίζονται οι πρωτεΐνες ενός βιολογικού δείγματος και αποκαλύπτονται τυχόν διαφοροποιήσεις στην έκφραση των πρωτεϊνών σε ένα συγκεκριμένο βιολογικό σύστημα. Οι πρωτεΐνες διαχωρίζονται σύμφωνα με το φορτίο τους και το μέγεθος τους σε ξεχωριστές κηλίδες. Η τοποθεσία όπου η πρωτεΐνη μεταναστεύει κατά την διάρκεια του διαχωρισμού είναι χαρακτηριστική για κάθε

συγκεκριμένη πρωτεΐνη, ενώ το μέγεθος και η ένταση της κηλίδας εξαρτώνται από την ποσότητα της πρωτεΐνης (Berth M., Moser F., Kolbe M., Bernhardt J., 2007).

Το αποτέλεσμα της δισδιάστατης ηλεκτροφόρησης πηκτωμάτων είναι μια ψηφιακή εικόνα, η οποία μπορεί να περιέχει μέχρι 10,000 πρωτεϊνικές κηλίδες, διάσπαρτα τοποθετημένες σε ανομοιογενές υπόβαθρο χαμηλής φωτεινότητας. Κάθε πρωτεϊνική κηλίδα έχει φωτεινότητα, σχήμα και μέγεθος. Γειτνιάζουσες πρωτεΐνες μπορεί να δημιουργούν συστάδες επικαλυπτόμενων πρωτεϊνών, ενώ υπολείμματα από πρωτεϊνικές κηλίδες μπορεί να δημιουργήσουν γραμμές στην εικόνα (streaks). Επιπλέον πρόβλημα στις εικόνες πρωτεομικής αποτελεί η ύπαρξη θορύβου καθώς και τεχνουργημάτων (artifacts) που οφείλονται σε ατέλειες του πειράματος, όπως σκόνη, γρατζουνιές κ.α (Dowsey A.W., Dunn M.J., Yang G.Z., 2003). Η κατάτμηση των εικόνων πρωτεομικής, παραμένει ακόμα και σήμερα επίπονη και χρονοβόρα διαδικασία για τους βιολόγους λόγω της προαναφερθείσας φύσης των εικόνων. Οι επιστήμονες της βιοπληροφορικής καλούνται να συμβάλουν στην επίλυση του προβλήματος, αναπτύσσοντας τεχνικές και λογισμικά που επιτυγχάνουν ικανοποιητικά αποτελέσματα. Οι υπάρχουσες τεχνικές και τα υπάρχοντα λογισμικά πακέτα, αδυνατούν να αντιμετωπίσουν με επιτυχία τα παραπάνω προβλήματα. Επιπρόσθετα, απαιτούν την επίπονη και χρονοβόρα παρέμβαση του χρήστη η οποία μειώνει την αντικειμενικότητα των αποτελεσμάτων.

Μέχρι στιγμής έχουν προταθεί αρκετές τεχνικές για την ανάλυση εικόνων πηκτώματος. Οι τεχνικές κατάτμησης εικόνων βασισμένες σε υδροκριτικές γραμμές (watersheds) (Pleissner K.P., Hoffman F., Kriegel K., Wenk C., Wegner S., 1999) μπορούν να οδηγήσουν σε υπερκατάτμηση. Η επιλογή εσωτερικών δεικτών (inner-markers) (Kim Y., Kim J., Won Y., In Y, 2003) είναι μια μέθοδος που προτάθηκε για την αντιμετώπιση του προβλήματος της υπερκατάτμησης, απαιτεί όμως ανθρώπινη παρέμβαση η οποία επηρεάζει τα αποτελέσματα. Οι μέθοδοι οι οποίες βασίζονται σε μορφολογία (morphology) (Vincent L., 1993) δεν δίνουν ικανοποιητικά αποτελέσματα λόγω της χαμηλής ποιότητας των εικόνων πηκτώματος. Μέθοδοι κατάτμησης εικόνων βασισμένες σε μονοδιάστατο ιστόγραμμα (1d histogram threshold) (Cutler P., Heald G., White I.R., Ruan J., 2003), αποτυγχάνουν γιατί το μονοδιάστατο ιστόγραμμα δεν περιέχει πληροφορία για τη χωρική συσχέτιση μεταξύ των εικονοστοιχείων της εικόνας. Κατ'επέκταση εικόνες με χαμηλό δείκτη σήματος-προς-θόρυβο (SNR) -όπως οι εικόνες πηκτώματος- δεν μπορούν να καταταμηθούν επιτυχώς με αυτήν την μέθοδο. Η μέθοδος Scimo είναι ανθεκτική στο θόρυβο, αλλά δεν καταφέρνει να εντοπίσει κηλίδες χαμηλής φωτεινότητας (Anjos A dos, Møller AL, Ersbøll BK, Finnie C., Shahbazkia H.R., 2010). Ανάμεσα στα διαθέσιμα λογισμικά πακέτα για ανάλυση εικόνων πηκτώματος είναι το Melanie 7 ([www.genebio.com](http://www.genebio.com)) και το Delta 2D ([www.decodon.com](http://www.decodon.com)), τα οποία διαθέτουν φιλική επαφή για τον χρήστη αλλά παρέχουν αμφιλεγόμενα αποτελέσματα διότι απαιτούν την επίπονη και χρονοβόρα παρέμβαση του χρήστη για την ρύθμιση παραμέτρων και την βελτιστοποίηση των αποτελεσμάτων.

Στην παρούσα εργασία παρουσιάζουμε την αναπτυχθείσα από την ερευνητική μας ομάδα αυτόματη μέθοδο κατάτμησης εικόνων πρωτεομικής, η οποία χρησιμοποιεί δισδιάστατα ιστογράμματα, τεχνικές κατωφλίωσης καθώς και υδροκριτικές γραμμές (watersheds). Η μέθοδος εφαρμόστηκε σε σύνολο εικόνων πρωτεομικής που περιέχουν πολλές επικαλυπτόμενες κηλίδες ποικίλων φωτεινοτήτων μεγεθών και σχημάτων, σε ανομοιογενές και θορυβώδες υπόβαθρο και τα αποτελέσματα της είναι καλύτερα από τα αποτελέσματα υπάρχοντων λογισμικών πακέτων καθώς και της μεθόδου Scimo. Η ακριβής κατάτμηση των πρωτεϊνικών κηλίδων παρουσιάζει εξαιρετικό ενδιαφέρον για τον κλάδο της υγείας γιατί μπορεί να οδηγήσει στην ανίχνευση διαγνωστικών δεικτών σε παθολογική και φυσιολογική κατάσταση και συνεπώς να αποκαλύψει ασθένειες που έχει ή πρόκειται να εκδηλώσει ο υπό εξέταση οργανισμός. Επιπρόσθετα μπορεί να συμβάλει σημαντικά στην εξατομικευμένη θεραπεία κάθε ασθενή.

## Υλικά και Μέθοδοι

Η προτεινόμενη μέθοδος έχει τα ακόλουθα τρία στάδια: Α) Εντοπισμός περιοχών που περιέχουν κυρίως κηλίδες, βάσει δισδιάστατων ιστογραμμάτων, Β) Εντοπισμός των κέντρων των κηλίδων, Γ) Κατάτμηση κηλίδων και εύρεση του ακριβούς τους περιγράμματος.

### Α. Εντοπισμός περιοχών που περιέχουν κυρίως κηλίδες, βάσει δισδιάστατων ιστογραμμάτων

Σε αυτό το στάδιο εντοπίζονται οι περιοχές της εικόνας στις οποίες υπάρχουν κυρίως κηλίδες. Πιο συγκεκριμένα η εικόνα κατατέμνεται χονδρικά σε ένα σύνολο από περιοχές που περιέχουν κυρίως υπόβαθρο  $B$  και σε ένα σύνολο από περιοχές οι οποίες περιέχουν κυρίως κηλίδες (σύνολο  $R$ ). Ο εντοπισμός των περιοχών που περιέχουν κηλίδες γίνεται σε δύο βήματα, ώστε να εντοπιστούν ακόμα και οι κηλίδες χαμηλής φωτεινότητας που είναι δύσκολα ορατές.

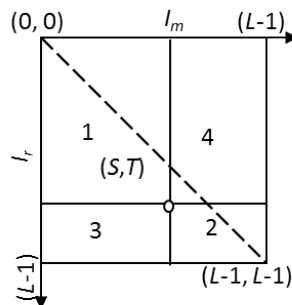
Τα δισδιάστατα ιστογράμματα στα οποία βασίζεται ο εντοπισμός των περιοχών που περιέχουν κυρίως κηλίδες παρουσιάζονται στην επόμενη παράγραφο, ενώ στο τέλος της παρούσας υποενότητας περιγράφονται λεπτομερώς τα δύο βήματα του εντοπισμού.

Έστω  $I$  μια εικόνα πηκτώματος και  $I_{mean}$  η εικόνα με τις μέσες τιμές της εικόνας  $I$  (Zhang J., Hu J., 2008). Το δισδιάστατο ιστόγραμμα απεικονίζεται στην Εικόνα 1. Ο  $x$ -άξονας και ο  $y$ -άξονας αντιπροσωπεύουν το επίπεδο φωτεινότητας των  $I$  και  $I_{mean}$  αντίστοιχα. Ο  $z$ -άξονας αντιπροσωπεύει την από κοινού συνάρτηση μάζας πιθανότητας (joint-probability mass function)  $P(I_r, I_m)$  η οποία δίνεται από την παρακάτω σχέση:

$$P(I_r, I_m) = \frac{r(I_r, I_m)}{M \times N} \quad (1)$$

Τα  $I_r$  και  $I_m$  αντιπροσωπεύουν τις τιμές φωτεινότητας δύο αντιστοιχών εικοστοιχείων που βρίσκονται στις εικόνες  $I$  και  $I_{mean}$ . Το  $r(I_r, I_m)$  αντιπροσωπεύει την συχνότητα εμφάνισης του ζεύγους  $(I_r, I_m)$ , ενώ τα  $M, N$  είναι οι διαστάσεις των δύο εικόνων  $I$  και  $I_{mean}$ . Σύμφωνα με την μέθοδο κατωφλίωσης 2D Otsu (Zhang J., Hu J., 2008) μπορεί να βρεθεί ένα ιδανικό διάνυσμα  $(S, T)$  το οποίο διασπά το δισδιάστατο ιστόγραμμα σε 4 τεταρτημόρια (Εικόνα 1). Τα τεταρτημόρια 1 και 2 περιέχουν τις κατανομές του υποβάθρου και του προσκήνιου ενώ τα τεταρτημόρια 3 και 4 περιέχουν τις κατανομές των εικονοστοιχείων που βρίσκονται κοντά σε ακμές και θόρυβο. Τα εικονοστοιχεία  $p$  που βρίσκονται στο προσκήνιο  $F$  είναι όσα ικανοποιούν την παρακάτω σχέση:

$$F = \{p: I_r(p) > S \wedge I_m(p) > T\} \quad (2)$$



Εικόνα 1: Δισδιάστατο Ιστογράμμα

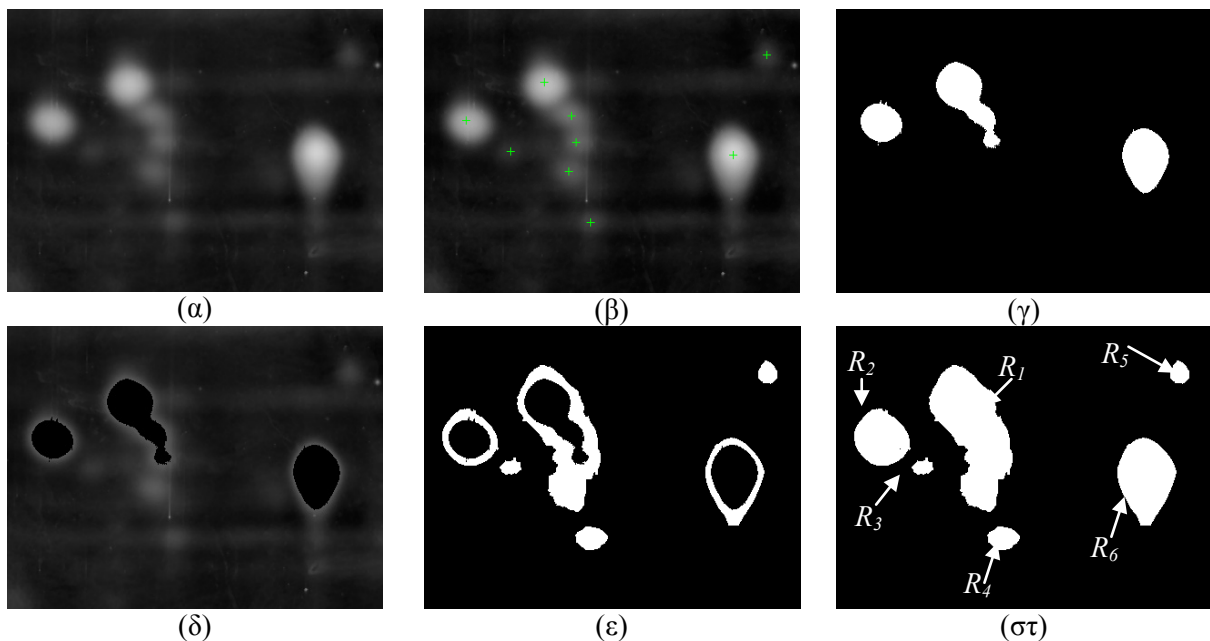
Στην παρακάτω εικόνα παρουσιάζονται τα δύο βήματα εντοπισμού των περιοχών που περιέχουν κηλίδες. Η αρχική Εικόνα 2α περιέχει 9 πρωτεΐνες, σύμφωνα με έμπειρους

βιολόγους της Βιοακαδημίας των Αθηνών, οι θέσεις των οποίων φαίνονται με πράσινους σταυρούς (ground truth) στην Εικόνα 2β. Για να εντοπιστούν οι 5 κηλίδες υψηλής φωτεινότητας της καθώς και οι υπόλοιπες 4 οι οποίες έχουν χαμηλή φωτεινότητα, η μέθοδος κατωφλίωσης 2D Otsu (Zhang J., Hu J., 2008) εφαρμόζεται δύο φορές. Αρχικά, εφαρμόζεται σε όλη την εικόνα, χωρίζοντας την σε προσκήνιο ( $F_0$ ) και υπόβαθρο ( $B_0$ ). Το προσκήνιο αποτελείται από ένα σύνολο περιοχών που περιλαμβάνουν κυρίως κηλίδες (άσπρες περιοχές Εικόνας 2γ). Τα υπόλοιπα εικονοστοιχεία (μαύρη περιοχή Εικόνας 2γ) είναι εκείνα που κατά πάσα πιθανότητα ανήκουν στο υπόβαθρο είτε σε κηλίδες χαμηλής φωτεινότητας ( $B_0$ ).

Για να εντοπιστούν κηλίδες που αποτελούνται από εικονοστοιχεία  $p$  χαμηλής φωτεινότητας, αλλά κατατάχθηκαν ως  $B_0$  στο προηγούμενο βήμα, η τεχνική κατωφλίωσης 2D Otsu εφαρμόζεται ξανά στα εικονοστοιχεία του υποβάθρου  $B_0$  (εικονοστοιχεία Εικόνας 2δ που δεν είναι μαύρα). Με αυτόν τον τρόπο η περιοχή  $B_0$  διαχωρίζεται στις περιοχές που δημιουργούν το  $F_{01}$  (άσπρες περιοχές Εικόνας 2ε) και στην περιοχή του υποβάθρου  $B_{01}$  (μαύρα εικονοστοιχεία Εικόνας 2ε που είναι μαύρα και στην Εικόνα 2γ). Το σύνολο των εικονοστοιχείων  $S$  που θεωρείται πως ανήκει σε κηλίδες ορίζεται ως:

$$S = \{p \mid p \in F_0 \cup F_{01}\} \quad (3)$$

και εμφανίζεται στην Εικόνα 2στ με άσπρο χρώμα. Στην ίδια εικόνα παρουσιάζονται οι 6 ανεξάρτητες συνεκτικές περιοχές  $R_i$ ,  $i=1..6$  που δημιουργούνται από τα εικονοστοιχεία του συνόλου  $S$ . Σύμφωνα με το ground truth η περιοχή  $R_1$  περιέχει 4 κηλίδες, ενώ καθεμία από τις περιοχές  $R_2-R_6$  περιέχει μία κηλίδα.

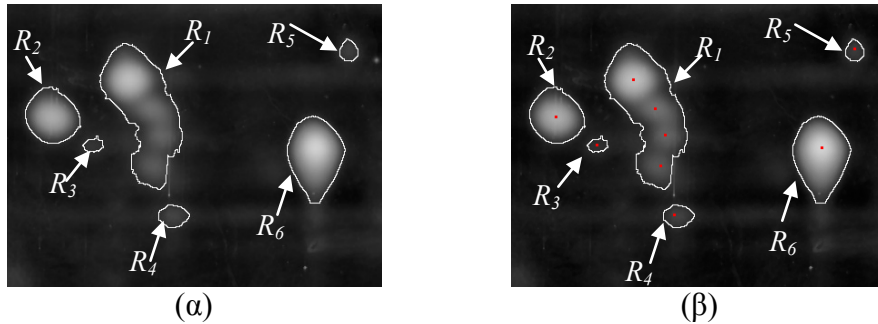


**Εικόνα 2.** Εντοπισμός των περιοχών που περιέχουν πρωτεΐνες σε δύο βήματα. (α) Αρχική εικόνα πηκτώματος, (β) θέσεις των κηλίδων (ground truth), (γ) κηλίδες υψηλής φωτεινότητας, (δ) κηλίδες χαμηλής φωτεινότητας και υπόβαθρο, (ε) κηλίδες χαμηλής φωτεινότητας, (στ) Το σύνολο των περιοχών που περιλαμβάνουν κυρίως κηλίδες ( $R_1$ -  $R_6$ ).

### B. Εντοπισμός κέντρων κηλίδων

Σκοπός αυτού του σταδίου είναι ο εντοπισμός των κέντρων των κηλίδων. Η φωτεινότητα των πρωτεϊνικών κηλίδων μεγιστοποιείται στο κέντρο τους ενώ μειώνεται σε περιοχές μακριά από το κέντρο τους (Bettens E., Scheunders P., Van Dyck D., Moens, L. Van Osta P., 1997).

Στις περιπτώσεις που η κορυφή είναι λεπτή οι κηλίδες μοιάζουν με την συνάρτηση 3D-Gaussian, ενώ στις περιπτώσεις που η κορυφή είναι πλατειά, οι κηλίδες μοιάζουν με οροπέδιο. Έτσι τα τοπικά μέγιστα στις περιοχές  $R_i$  είναι οι πιθανότεροι υποψήφιοι για τα κέντρα των κηλίδων που υπάρχουν σε αυτές. Στην Εικόνα 3α παρουσιάζονται τα περιγράμματα των περιοχών που περιέχουν κηλίδες όπως αυτά βρέθηκαν από το προηγούμενο στάδιο και στην Εικόνα 3β παρουσιάζονται με κόκκινο τα κέντρα των 9 κηλίδων που βρίσκονται στις περιοχές  $R_1$ - $R_6$  όπως αυτά εντοπίστηκαν από αυτό το στάδιο.

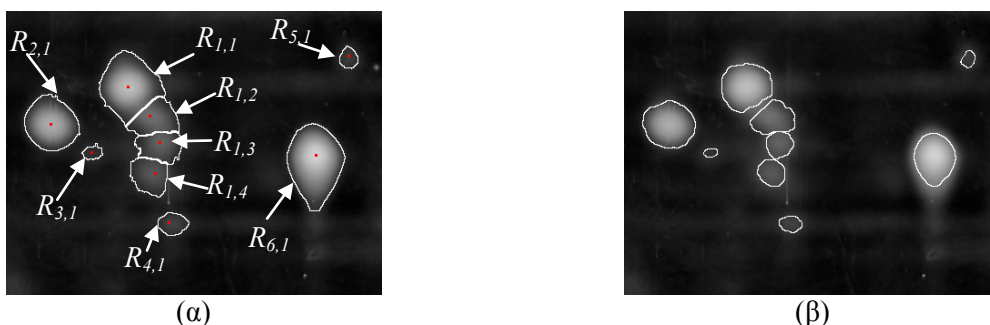


Εικόνα 3. (α) Περιοχές που περιέχουν κυρίως κηλίδες, (β) κέντρα κηλίδων.

#### Γ. Κατάτμηση κηλίδων και εύρεση του ακριβούς περιγράμματος τους

Σε αυτό το στάδιο κάθε περιοχή  $R_i$  διαχωρίζεται στις επικαλυπτόμενες υποπεριοχές  $R_{ij}$  από τις οποίες αποτελείται και εν συνεχεία εντοπίζεται το ακριβές περίγραμμα αυτών. Η κατάτμηση κάθε περιοχής  $R_i$  γίνεται με βάση τα κέντρα των κηλίδων που εντοπίστηκαν στο προηγούμενο στάδιο κάνοντας χρήση υδροκριτικών γραμμών (watersheds). Αρχικά δίνεται μία ετικέτα σε κάθε τοπικό μέγιστο της  $R_i$  η οποία δηλώνει την υποπεριοχή  $j$  την οποία αρχικοποιεί. Στην συνέχεια ελέγχεται κάθε εικονοστοιχείο  $p$  της  $R_i$  κατά φθίνουσα σειρά φωτεινότητας  $I$ . Κάθε εικονοστοιχείο  $p$  λαμβάνει την ετικέτα του εικονοστοιχείου με το οποίο γειτνιάζει και εκχωρείται με αυτόν τον τρόπο στην υποπεριοχή αυτή. Με τον τρόπο αυτό οι 6 περιοχές  $R_i$  χωρίζονται σε 9 υποπεριοχές, όσες δηλαδή είναι και οι ετικέτες με τις οποίες επισημάνθηκαν τα εικονοστοιχεία του  $S$  με την διαδικασία αυτή. Είναι προφανές πως τα εικονοστοιχεία που βρίσκονται σε κάθε υποπεριοχή έχουν ίδια τιμή-ετικέτα ενώ τα εικονοστοιχεία διαφορετικών υποπεριοχών έχουν διαφορετικές μεταξύ τους ετικέτες. Η Εικόνα 4α παρουσιάζει τα περιγράμματα των κηλίδων. Κάθε μία από αυτές τις υποπεριοχές θεωρείται κηλίδα η οποία πιθανόν να περιλαμβάνει και υπόβαθρο.

Για να αφαιρεθούν τα εικονοστοιχεία του συνόλου  $S$  που βρίσκονται στις υποπεριοχές  $R_{ij}$  και ανήκουν σε υπόβαθρο υψηλής φωτεινότητας, εφαρμόζεται σε κάθε  $R_{ij}$  ο αλγόριθμος βέλτιστου κατωφλιού (Gonzalez R., Woods R., 2007). Στην Εικόνα 4β παρουσιάζονται τα τελικά περιγράμματα της μεθόδου μας. Όπως φαίνεται η μέθοδος μας εντόπισε και κατέτμησε επιτυχώς τις 9 πρωτείνες της Εικόνας 2α.



Εικόνα 4. (α) Κατετμημένες περιοχές κηλίδων, (β) ακριβή περιγράμματα κηλίδων.

## Αποτελέσματα και Συζήτηση

Η προτεινόμενη μέθοδος εφαρμόστηκε σε πραγματικές εικόνες πηκτώματος οι οποίες περιείχαν περίπου 10.000 κηλίδες -ανάμεσα τους και πολλές επικαλυπτόμενες- ποικίλων φωτεινοτήτων, μεγεθών και σχημάτων σε ανομοιογενές και θορυβώδες υπόβαθρο. Οι εικόνες αυτές καθώς και οι θέσεις των κηλίδων σε αυτές παραχωρήθηκαν ευγενώς από έμπειρους βιολόγους της Βιοακαδημίας Αθηνών.

Τα αποτελέσματα που προέκυψαν από τα λογισμικά πακέτα Melanie 7, Delta 2D καθώς και την μέθοδο Scimo καθώς και την μέθοδό μας, αξιολογήθηκαν ως προς την επιτυχία εντοπισμού και κατάτμησης των κηλίδων χρησιμοποιώντας το ground truth των έμπειρων βιολόγων και παρουσιάζονται στον Πίνακα 1.

Η μέθοδος μας εντόπισε το 79,3% των πραγματικών κηλίδων, ενώ το Melanie 7, το Delta 2D και η Scimo μόλις το 65,9%, 63,2% και 60,3%, αντίστοιχα. Επιπρόσθετα το ποσοστό των ψευδών κηλίδων που εντοπίστηκαν από την μέθοδο μας (3,0%) ήταν συγκριτικά χαμηλότερο από αυτό των Melanie 7 (5,0%), το Delta 2D (20,6%) και της Scimo (8,1%).

Ένα παράδειγμα περιοχής εικόνας πηκτώματος παρουσιάζεται στην Εικόνα 5α. Παρόλο που το μέγεθος της είναι μικρό περιλαμβάνει 132 κηλίδες διαφόρων φωτεινοτήτων, μεγεθών και σχημάτων. Οι θέσεις των κηλίδων, πολλές από τις οποίες επικαλύπτονται, παρουσιάζονται με πράσινους σταυρούς στις Εικόνες 5β έως και 5στ. Τα αποτελέσματα εντοπισμού και κατάτμησης της Εικόνας 5α, τα οποία προέκυψαν από τα λογισμικά πακέτα Melanie 7 και Delta 2D, παρουσιάζονται στις Εικόνες 5γ και 5δ, ενώ τα αποτελέσματα που προέκυψαν από την μέθοδο Scimo καθώς και τη μέθοδό μας παρουσιάζονται στις Εικόνες 5ε και 5στ. Σταυροί που δεν εσωκλείονται σε περίγραμμα δηλώνουν κηλίδες που δεν εντοπίστηκαν και δεν κατατμήστηκαν, ενώ περιγράμματα που δεν περιλαμβάνουν σταυρό δηλώνουν ψευδώς θετικές κηλίδες, δηλαδή περιγράμματα που δεν αντιστοιχούν σε πραγματικές κηλίδες.

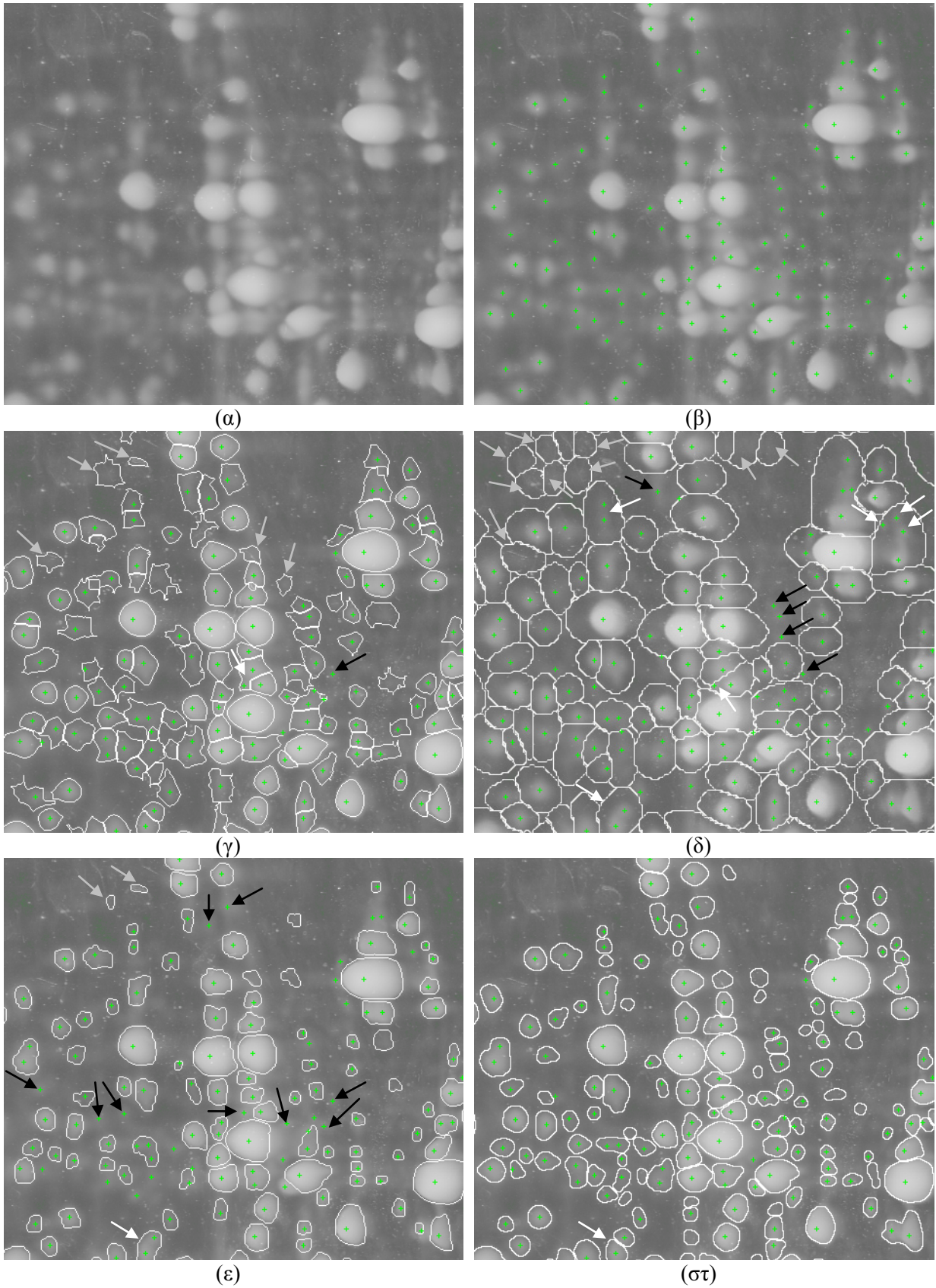
Το λογισμικό πακέτο Melanie 7 εντόπισε 121 από τις 132 πραγματικές κηλίδες ενώ βρήκε 17 ψευδείς κηλίδες (Εικόνα 5γ). Το Delta 2D εντόπισε μόνο τις 95 κηλίδες και βρήκε επιπρόσθετα 11 ψευδείς κηλίδες (Εικόνα 5δ). Η μέθοδος Scimo εντόπισε τις 101 από τις 132 κηλίδες ενώ βρήκε επιπρόσθετα 7 ψευδείς κηλίδες (Εικόνα 5ε). Η μέθοδος μας εντόπισε το μεγαλύτερο πλήθος πραγματικών κηλίδων συγκριτικά με όλες τις άλλες μεθόδους (123 από τις 132) ενώ βρήκε επιπρόσθετα 16 ψευδείς κηλίδες (Εικόνα 5στ).

Στις εικόνες 5γ-5ε παρουσιάζονται ενδεικτικά με μαύρα βέλη παραδείγματα κηλίδων που το Melanie 7, το Delta 2D και το Scimo, αντίστοιχα, δεν εντόπισαν, ενώ παραδείγματα ψευδών κηλίδων τις οποίες το Melanie 7, το Delta 2D και η μέθοδος Scimo λανθασμένα εντόπισαν και κατέτμησαν παρουσιάζονται με γκρι βέλη. Τέλος στις ίδιες εικόνες, με λευκά βέλη παρουσιάζονται επικαλυπτόμενες κηλίδες που δεν διαχωρίστηκαν. Παρατηρώντας τα αποτελέσματα της μεθόδου μας (Εικόνα 5στ), στα αντίστοιχα σημεία, βλέπουμε πως η προτεινόμενη μέθοδος έχει καταφέρει να εντοπίσει και κατατμήσει επιτυχώς όλες αυτές τις κηλίδες, παρέχοντας συγκριτικά καλύτερα αποτελέσματα από τα αντίστοιχα των άλλων προγραμμάτων και μεθόδων.

	Melanie 7	Delta 2D	Scimo	Μέθοδός Μας
Αληθώς Εντοπισμένες κηλίδες	65,9%	63,2%	60,3%	79,3%
Ψευδώς Θετικές Κηλίδες	5,0%	20,6%	8,1%	3,0%

Πίνακας 1. Αποτελέσματα εντοπισμού κηλίδων





**Εικόνα 5.** (α) Αρχική εικόνα, (β) θέσεις κηλίδων (ground truth), αποτελέσματα εντοπισμού και κατάτμησης (γ) MelanIE 7, (δ) Delta 2D, (ε) Scimo, (στ) Μεθόδου μας.

## Συμπεράσματα

Συνοψίζοντας, η δισδιάστατη ηλεκτροφόρηση πηκτώματος είναι μια ισχυρή, καθιερωμένη μέθοδος για τον διαχωρισμό χιλιάδων πρωτεϊνών μέσα σε ένα βιολογικό δείγμα. Ο εντοπισμός και ο ακριβής προσδιορισμός της ποσότητας κάθε πρωτεΐνης στα κύτταρα ενός οργανισμού, ο οποίος επιτυγχάνεται μέσω του ακριβούς προσδιορισμού του περιγράμματός της, παρουσιάζει εξαιρετικό ενδιαφέρον, γιατί η διαφοροποίηση της ποσότητας αυτής σχετίζεται άμεσα με τις ασθένειες που έχει ή πρόκειται να εκδηλώσει ο υπό εξέταση οργανισμός. Η ανάλυση των εικόνων πηκτώματος απαιτεί την χρήση αποτελεσματικών μεθόδων επεξεργασίας εικόνας ώστε να εντοπιστούν και διαχωριστούν οι πρωτεΐνες που υπάρχουν σε αυτές. Η επιτυχής Τα διαθέσιμα λογισμικά ανάλυσης εικόνων πηκτώματος δεν παρέχουν ικανοποιητικά αποτελέσματα και συνήθως απαιτούν την παρέμβαση του χρήστη για την βελτιστοποίηση των αποτελεσμάτων. Στην παρούσα εργασία παρουσιάζεται η αναπτυχθείσα από την ομάδα μας πρωτότυπη μέθοδος ανάλυσης εικόνων πηκτώματος. Η μέθοδος αυτή βασίζεται στον εντοπισμό των περιοχών της εικόνας που περιέχουν πρωτεϊνικές κηλίδες χρησιμοποιώντας δισδιάστατα ιστογραμμάτα. Εν συνεχεία, τα κέντρα των κηλίδων που υπάρχουν σε αυτές εντοπίζονται και οι περιοχές αυτές διαχωρίζονται γύρω από τα κέντρα των πρωτεϊνών με χρήση watersheds. Τέλος, εφαρμόζεται ο αλγόριθμος βέλτιστης κατωφλίωσης σε κάθε υποπεριοχή για την απομάκρυνση του υποβάθρου του περιστοιχίζει κάθε κηλίδα και την εύρεση του τελικού περιγράμματος.

Η πειραματική αξιολόγηση που διενεργήθηκε απέδειξε πως η συγκεκριμένη μέθοδος εντοπίζει και κατατμεί με υψηλά ποσοστά επιτυχίας την πλειοψηφία των πρωτεϊνικών κηλίδων που υπάρχουν σε πραγματικές εικόνες πηκτώματος. Τα αποτελέσματα που προέκυψαν από την μεθόδό μας είναι συγκριτικά καλύτερα από αυτά των λογισμικών πακέτων Melanie 7 και Delta 2D καθώς και της μεθόδου Scimo.

## Ευχαριστίες

Θα θέλαμε να ευχαριστήσουμε το Ίδρυμα Ιατροβιολογικών Ερευνών της Ακαδημίας Αθηνών (Ι.ΙΒ.Ε.Α.Α.) για την παροχή των εικόνων πηκτώματος καθώς και του ground truth. Η παρούσα εργασία χρηματοδοτήθηκε από το Ίδρυμα Κρατικών Υποτροφιών (Ι.Κ.Υ.).

## Βιβλιογραφία

Aamodt R., Bondi J., Andersen S.N., Bakka A., Bukholm G., Bukholm I., (2010), *The Prognostic Impact of Protein Expression of E-Cadherin-Catenin Complexes Differs between Rectal and Colon Carcinoma*, Gastroenterology Research and Practice (doi:10.1155/2010/616023)

Anjos A dos, Møller AL, Ersbøll BK, Finnie C., Shahbazkia H.R., (2010), *New approach for Segmentation and Quantification of Two-Dimensional Gel Electrophoresis Images*, Bioinformatics, vol. 27, no.3, pp. 368-375

Berth M., Moser F., Kolbe M., Bernhardt J., (2007), *The state of the art in the analysis of two-dimensional gel electrophoresis images*, Applied Microbiology and Biotechnology, no. 76, pp. 1223-1243

Bettens E., Scheunders P., Van Dyck D., Moens, L. Van Osta P., (1997), *Computer analysis of two-dimensional electrophoresis gels: a new segmentation and modelling algorithm*, Electrophoresis, vol. 18, pp. 792-798

Cutler P., G. Heald G., White I.R., Ruan J., (2003), *A novel approach to spot detection for two-dimensional gel electrophoresis images using pixel value collection*, Proteomics, vol. 3, pp. 392-401

Delta2D ([www.decodon.com](http://www.decodon.com))



Dowsey A.W., Dunn M.J., Yang G.Z., (2003), *The role of bioinformatics in two-dimensional gel electrophoresis*, Proteomics J., vol. 3, no. 8, pp. 1567–1596

Gonzalez R., Woods R., (2007), *Digital Image Processing*, Prentice Hall, New Jersey.

Kim Y., Kim J., Won, Y. , In Y.,(2003), *Segmentation of protein spots in 2-D gel electrophoresis images with watershed using hierarchical threshold*, LNCS, vol. 2869, pp. 389-396

Melanie 7 (www.genebio.com)

Pleissner K.P., Hoffman F., Kriegel K., Wenk C., Wegner S., Sahlstrom A., Oswald H., Alt H., Fleck E., (1999), *New algorithmic approaches to protein spot detection and pattern matching in two-dimensional electrophoresis databases*, Electrophoresis, vol. 20, pp. 755-765

Roti G., Stegmaier K., 2012, *Genetic and proteomic approaches to identify cancer drug targets*, *British Journal of Cancer*, vol. 106, pp. 254–261

Ruckhäberlea E., Karna T., Hankera L., Schwarzb J., Schulz-Knappeb P., Kuhnb K., Böhm G., Selzerb S., Neukumb E., Engelse K., Holtricha U., Kaufmanna M., Rodya A., (2010) *Differences in Protein Expression between Estrogen Receptor-Positive and -Negative Tumors Identified by Tandem Mass Tag Technology*, *Breast Cancer Proteomics*, vol. 5, pp. 5-7

Stannard C., Brown L.R., Godovac-Zimmermann J., (2004), *New Paradigms in Cellular Function and the Need for Top-Down Proteomics Analysis*, *Current Proteomics*, vol. 1, pp. 13-24

Vincent L., Soille P.Q., (1991), *Watersheds in digital spaces: an efficient algorithm based on immersion simulations*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 583–598

Vincent L., (1993), *Morphological grayscale reconstruction in image analysis: applications and efficient algorithms*, *IEEE Trans. on Image Processing*, vol. 2, no. 2, pp. 176-201

Zhang J., Hu, J., (2008), *Image Segmentation Based on 2D Otsu Method with Histogram Analysis*, *International Conference Science and Software Engineering*