



ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

ΙΑΤΡΙΚΗ ΣΧΟΛΗ

ΛΕΙΤΟΥΡΓΙΚΟΣ-ΚΛΙΝΙΚΟΕΡΓΑΣΤΗΡΙΑΚΟΣ ΤΟΜΕΑΣ

ΕΡΓΑΣΤΗΡΙΟ ΒΙΟΛΟΓΙΚΗΣ ΧΗΜΕΙΑΣ

**«Επεξεργασία Εικόνων Μικροσυστοιχιών με Χρήση
Ευφρών Μεθόδων Πληροφοριακών Συστημάτων»**

ΝΙΚΟΛΑΟΣ ΓΙΑΝΝΑΚΕΑΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΙΩΑΝΝΙΝΑ 2011



ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

ΙΑΤΡΙΚΗ ΣΧΟΛΗ

ΛΕΙΤΟΥΡΓΙΚΟΣ-ΚΛΙΝΙΚΟΕΡΓΑΣΤΗΡΙΑΚΟΣ ΤΟΜΕΑΣ

ΕΡΓΑΣΤΗΡΙΟ ΒΙΟΛΟΓΙΚΗΣ ΧΗΜΕΙΑΣ

**«Επεξεργασία Εικόνων Μικροσυστοιχιών με Χρήση
Ευφών Πληροφοριακών Συστημάτων»**

ΝΙΚΟΛΑΟΣ ΓΙΑΝΝΑΚΕΑΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΙΩΑΝΝΙΝΑ 2011

«Η έγκριση της διδακτορικής διατριβής από την Ιατρική Σχολή του Πανεπιστημίου Ιωαννίνων δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα»

Ν. 5343/32, άρθρο 202, παράγραφος 2

Εδώ μπαίνει το χαρτί από την γραμματεία

ΕΥΧΑΡΙΣΤΙΕΣ

Η εργασία πραγματοποιήθηκε στα πλαίσια της διδακτορικής μου διατριβής υπό την επίβλεψη του κ. Θεόδωρου Φώτση, Καθηγητή της Ιατρικής Σχολής του Πανεπιστημίου Ιωαννίνων. Μέλη της τριμελούς συμβουλευτικής επιτροπής ήταν ο κ. Δημήτριος Φωτιάδης, Καθηγητής του Τμήματος Μηχανικών Επιστήμης Υλικών και η κ. Αναστασία Πολίτου, Επίκουρη καθηγήτρια της Ιατρικής Σχολής.

Θα ήθελα να ευχαριστήσω ιδιαίτερα τον καθ. Δ. Φωτιάδη που με υποστήριξε υλικά επιστημονικά αλλά κυρίως ηθικά και πνευματικά. Με την επιστημονική υποστήριξη που μου παρείχε κατέστη δυνατή η εκπόνηση της διδακτορικής διατριβής, από την οποία παρείχθησαν 4 δημοσιεύσεις σε διεθνή επιστημονικά περιοδικά, 12 εργασίες σε πρακτικά συνεδρίων και 2 κεφάλαια σε βιβλία. Επίσης η οικονομική υποστήριξη από την Μονάδα Ιατρικής Τεχνολογίας και Ευφών Πληροφοριακών Συστημάτων υπήρξε πολύ σημαντικός αρωγός της προσπάθειας μου, καθώς εκτός των άλλων μου παρείχε την δυνατότητα συμμετοχής στα σημαντικότερα επιστημονικά συνέδρια στον τομέα μου.

Οφείλω τέλος ένα μεγάλο ευχαριστώ στο οικογενειακό μου περιβάλλον που με στήριξε ηθικά αυτά τα χρόνια. Τους φίλους μου και συνεργάτες από την Μονάδα Ιατρικής Τεχνολογίας και Ευφών Πληροφοριακών Συστημάτων, Θέμη Έξαρχο, Πέτρο Καρβέλη, Αλέξανδρο Τζάλλα, Φάνη Καλατζή, Αναστάσιο Κούτλα, Μάρκο Τσίπουρα, Δημήτριο Τσαλικάκη, Ευάγγελο Καρβούνη και Αντώνιο Σακελάριο.

ΠΕΡΙΕΧΟΜΕΝΑ

| | |
|---|----|
| Κεφάλαιο 1 ^ο : Βιολογικό Υπόβαθρο | 1 |
| 1.1. Βασικές Έννοιες | 1 |
| 1.2. Η δομή και η λειτουργία του DNA | 3 |
| 1.3. Αντιγραφή του DNA | 8 |
| 1.4. Μεταλλάξεις και Απλοί Νουκλεοτιδικοί Πολυμορφισμοί..... | 14 |
| 1.5. Γονιδιακή Έκφραση | 15 |
| 1.5.1. Μεταγραφή του DNA σε RNA | 16 |
| 1.5.2. Αντίστροφη μεταγραφή..... | 20 |
| 1.5.3. Μετάφραση | 21 |
| 1.6. Υβριδοποίηση νουκλεϊκών οξέων..... | 22 |
| 1.7. Αλυσιδωτή Αντίδραση Πολυμεράσης | 24 |
| 1.8. Συμπεράσματα..... | 26 |
| ΚΕΦΑΛΑΙΟ 2 ^ο : Εισαγωγή στις Μικροσυστοιχίες | 27 |
| 2. Εισαγωγή στις μικροσυστοιχίες | 27 |
| 2.1. Ιστορική αναδρομή | 27 |
| 2.2. Προετοιμασία του Πειράματος και βιολογικό πείραμα | 31 |
| 2.2.1. Βιολογικό Ερώτημα | 32 |
| 2.2.2. Εκτύπωση των ανιχνευτών..... | 34 |
| 2.2.2.1. Εκτύπωση με επαφή..... | 36 |
| 2.2.2.2. Εκτύπωση Χωρίς επαφή..... | 38 |
| 2.2.3. Υβριδοποίηση των δειγμάτων με τους ανιχνευτές..... | 40 |
| 2.2.4. Σάρωση..... | 41 |
| 2.3. Σύνολα Δεδομένων..... | 44 |
| 2.3.1. Βάση δεδομένων του Στάνφορντ | 45 |

| | | |
|---|---|----|
| 2.3.2. | Εικόνες από την μικροσυστοιχία CNV370 της Illumina | 46 |
| 2.3.3. | Προσομοιωμένες Εικόνες..... | 47 |
| 2.4. | Συμπεράσματα..... | 50 |
| ΚΕΦΑΛΑΙΟ 3 ^ο : Βιβλιογραφική ανασκόπηση..... | | 51 |
| 3. | Βιβλιογραφική ανασκόπηση | 51 |
| 3.1. | Εισαγωγή στις εικόνες μικροσυστοιχιών | 51 |
| 3.2. | Εντοπισμός των κηλίδων και τοποθέτηση πλέγματος..... | 54 |
| 3.2.1. | Βελτίωση της εικόνας..... | 58 |
| 3.2.2. | Διαχωρισμός των συστάδων..... | 61 |
| 3.2.3. | Ολιστική Προσέγγιση..... | 63 |
| 3.2.4. | Προσέγγιση Κηλίδα με κηλίδα..... | 64 |
| 3.2.5. | Τοποθέτηση Πλέγματος | 67 |
| 3.2.6. | Βελτίωση του πλέγματος..... | 68 |
| 3.3. | Κατάτμηση των κηλίδων..... | 69 |
| 3.3.1. | Μέθοδοι βασισμένες σε τεχνικές επεξεργασίας εικόνων | 70 |
| 3.3.1.1. | Μέθοδοι σταθερού ή μεταβλητού κύκλου | 71 |
| 3.3.1.2. | Μέθοδοι βασισμένες στο ιστόγραμμα της εικόνας | 74 |
| 3.3.1.3. | Μέθοδοι Μεταβλητού σχήματος..... | 76 |
| 3.3.2. | Μέθοδοι βασισμένες σε τεχνικές Μηχανικής Μάθησης..... | 79 |
| 3.3.2.1. | Κατάτμηση με τον αλγόριθμο K-μέσων..... | 79 |
| 3.3.2.2. | Κατάτμηση με τον ασαφή αλγόριθμο C-μέσων | 80 |
| 3.3.2.3. | Κατάτμηση με τυχαία Μαρκοβιανά πεδία | 81 |
| 3.3.2.4. | Κατάτμηση με διαμερισμό γύρω από το μέσο | 82 |
| 3.3.2.5. | Κατάτμηση με Μοντέλα Μεικτών Γκαουσιανών..... | 82 |
| 3.3.3. | Υβριδική Μέθοδος | 83 |

| | | |
|--|--|-----|
| 3.4. | Εξαγωγή Εντάσεων | 83 |
| 3.5. | Συμπεράσματα..... | 85 |
| Κεφάλαιο 4ο: Εντοπισμός Κηλίδων και τοποθέτηση πλέγματος | | 87 |
| 4. | Εντοπισμός Κηλίδων και τοποθέτηση πλέγματος..... | 87 |
| 4.1. | Εισαγωγή..... | 87 |
| 4.2. | Εύρεση των Συστάδων | 90 |
| 4.2.1. | Αρχική Προσέγγιση | 91 |
| 4.2.2. | Γενικευμένη προσέγγιση..... | 94 |
| 4.3. | Εντοπισμός των υβριδοποιημένων κηλίδων | 98 |
| 4.4. | Εντοπισμός των κηλίδων που δεν έχουν υβριδοποιηθεί | 105 |
| 4.4.1. | Εντοπισμός μη υβριδοποιημένων κηλίδων σε εικόνες τετραγωνικού πλέγματος | 105 |
| 4.4.2. | Γενικευμένη αλγόριθμος εντοπισμού των μη υβριδοποιημένων κηλίδων 108 | |
| 4.5. | Τοποθέτηση πλέγματος..... | 118 |
| 4.6. | Αποτελέσματα..... | 122 |
| 4.6.1. | Αρχική προσέγγιση τετραγωνικού πλέγματος | 123 |
| 4.6.2. | Εξαγωνικό Πλέγμα..... | 126 |
| 4.6.3. | Γενικευμένη Μέθοδος..... | 132 |
| 4.7. | Συμπεράσματα..... | 134 |
| Κεφάλαιο 5: Κατάτμηση εικόνων μικροσυστοιχιών | | 139 |
| 5. | Κατάτμηση εικόνων μικροσυστοιχιών..... | 139 |
| 5.1. | Εισαγωγή..... | 139 |
| 5.2. | Εξαγωγή των χαρακτηριστικών | 140 |
| 5.3. | Μείωση της διάστασης των χαρακτηριστικών με χρήση Ανάλυσης Κύριων Συνιστωσών..... | 145 |

| | | |
|--------------|--|-----|
| 5.4. | Διαχωρισμός Εικονοστοιχείων με χρήση τεχνικών Μηχανικής Μάθησης... | 147 |
| 5.4.1. | Κατάτμηση χωρίς επίβλεψη | 150 |
| 5.4.1.1. | Ομαδοποίηση με τον αλγόριθμο K-μέσων | 150 |
| 5.4.2. | Ομαδοποίηση με τον Ασαφή αλγόριθμο K-μέσων | 152 |
| 5.4.2.1. | Αντιστοίχιση ομάδων με κατηγορία..... | 153 |
| 5.4.3. | Κατάτμηση με επίβλεψη..... | 155 |
| 5.4.3.1. | Μπεϋζιανός Ταξινομητής..... | 156 |
| 5.4.3.2. | Μηχανές Διανυσμάτων Υποστήριξης | 157 |
| 5.5. | Σύνολα δεδομένων | 161 |
| 5.6. | Αποτελέσματα | 163 |
| 5.6.1. | Κατάτμηση χωρίς επίβλεψη | 163 |
| 5.6.2. | Κατάτμηση με επίβλεψη..... | 170 |
| 5.6.2.1. | Πραγματικές εικόνες | 170 |
| 5.6.2.2. | Προσομοιωμένες εικόνες..... | 172 |
| 5.7. | Συμπεράσματα..... | 175 |
| Κεφάλαιο 6°: | Συμπεράσματα | 179 |
| 6. | Συμπεράσματα..... | 179 |

ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

| | |
|---|----|
| Σχήμα 1.1: Η ύπαρξη των γονιδίων στα χρωμοσώματα των ζωντανών οργανισμών..... | 3 |
| Σχήμα 1.2: Η δίκλωνη αλυσίδα του DNA..... | 5 |
| Σχήμα 1.3: Η δομή του DNA | 6 |
| Σχήμα 1.4: Το Ριβονουκλεϊκό οξύ (RNA) σε σύγκριση με το δεοξυριβονουκλεϊκό οξύ (DNA)..... | 13 |
| Σχήμα 1.5: Η διαδικασία της αντιγραφής. | 14 |
| Σχήμα 1.6: Ένας απλός νουκλεοτιδικός πολυμορφισμός (SNP). | 15 |
| Σχήμα 1.7: Οι διαδικασίες με τις οποίες εκφράζεται η γενετική πληροφορία..... | 16 |
| Σχήμα 1.8: Ο μηχανισμός της μεταγραφής. | 20 |
| Σχήμα 1.9: Η διεργασία της υβριδοποίηση του DNA..... | 23 |
| Σχήμα 1.10: Η αλυσιδωτή Αντίδραση πολυμεράσης..... | 25 |
| Σχήμα 2.1: Ο κύκλος ζωής των μικροσυστοιχιών | 28 |
| Σχήμα 2.2: Η εξελικτική πορεία της τεχνολογίας των μικροσυστοιχιών. | 30 |
| Σχήμα 2.3: Η ραγδαία εξέλιξη της τεχνολογίας μικροσυστοιχιών όπως αποτυπώνεται από τον αριθμό των δημοσιεύσεων τα τελευταία χρόνια..... | 31 |
| Σχήμα 2.4: Τα στάδια του κύκλου ζωής των μικροσυστοιχιών τα οποία προηγούνται της επεξεργασία των εξαγόμενων εικόνων. | 32 |
| Σχήμα 2.5: Κατηγοριοποίηση των ειδών των μικροσυστοιχιών ανάλογα με της βιολογικές αλληλουχίες που χρησιμοποιούνται..... | 33 |
| Σχήμα 2.6: Η τεχνικές για την εκτύπωση μικροσυστοιχιών. | 35 |
| Σχήμα 2.7: Συσκευή εκτύπωσης μικροσυστοιχιών με ακίδες..... | 36 |
| Σχήμα 2.8: Διαφορετικοί σχεδιασμοί ακίδων: α) συμπαγής ακίδα β), γ), και δ) διαφορετικές ακίδες με σχισμή. | 37 |
| Σχήμα 2.9: Οι δομή των φθορίζουσών ουσιών Cy3 και Cy5..... | 40 |
| Σχήμα 2.10: Δύο πολύ γνωστοί σαρωτές μικροσυστοιχιών α) της εταιρίας Genepix και β) της εταιρίας Agilent. | 42 |

| | |
|---|----|
| Σχήμα 2.11: Τα δύο κανάλια στο αριστερό μέρος της εικόνας και οι έγχρωμη εικόνα μικροσυστοιχιών..... | 43 |
| Σχήμα 2.12: Εικόνα μικροσυστοιχιών με 16 συστάδες και 576 κηλίδες σε κάθε συστάδα | 44 |
| Σχήμα 2.13: Η μικροσυστοιχία Beadchip CNV370 της εταιρίας Illumina και η εξαγόμενη εικόνα. | 47 |
| Σχήμα 2.14: Διάγραμμα με τις μονάδες του προσομοιωτή του Nykter. | 48 |
| Σχήμα 3.1: Γενικό διάγραμμα ροής της επεξεργασίας των εικόνων των μικροσυστοιχιών | 54 |
| Σχήμα 3.2: Διάγραμμα ροής του εντοπισμού των κηλίδων | 55 |
| Σχήμα 3.3: Δενδρόγραμμα των μεθόδων για τον εντοπισμό των κηλίδων | 58 |
| Σχήμα 3.4: Το αποτέλεσμα των πολυκαναλικών φίλτρων σε τρεις διαφορετικές κηλίδες. 60 | |
| Σχήμα 3.5: Διόρθωση της στρέψης των εικόνων μικροσυστοιχιών υπολογίζοντας το φάσμα ισχύος τις εικόνας.: α) εικόνα χωρίς στρέψη, β) στραμμένη εικόνα. | 61 |
| Σχήμα 3.6: Η κατακόρυφη και η οριζόντια προβολή μιας εικόνας με 16 συστάδες. Η ευθείες που ορίζονται από της κοιλάδες των δύο προβολών χωρίζουν την εικόνα στις επιμέρους συστάδες, στα σημεία στα οποία τέμνονται..... | 62 |
| Σχήμα 3.7. Η οριζόντια προβολή μίας συστάδας..... | 63 |
| Σχήμα 3.8: Η τοποθέτηση πλέγματος σε μια εικόνα μικροσυστοιχιών. (α) Τοποθέτηση σταθερού κελιού σε κάθε μία κηλίδα, (β) εφαρμογή του διαγράμματος Βορονόι..... | 68 |
| Σχήμα 3.9: Δενδρόγραμμα των μεθόδων που έχουν παρουσιαστεί στην βιβλιογραφία για την κατάτμηση των εικόνων των μικροσυστοιχιών. | 72 |
| Σχήμα 3.10: Κατάτμηση με μέθοδο σταθερού κύκλου. | 73 |
| Σχήμα 3.11: Κατάτμηση με το πρόγραμμα Genepix..... | 74 |
| Σχήμα 3.12: Μέρος μιας εικόνας μικροσυστοιχιών θεωρούμενη ως ανάγλυφη τοποθεσία. | 78 |
| Σχήμα 3.13: Οι περιοχές από τις οποίες λαμβάνονταν τα εικονοστοιχεία υποβάθρου από τις πρώτες μεθόδους κατάτμησης..... | 85 |
| Σχήμα 4.1: Σύγκριση μεταξύ του τετραγωνικού και του εξαγωνικού πλέγματος..... | 88 |

| | |
|--|-----|
| Σχήμα 4.2: Τα τέσσερα βήματα σταδίου της εύρεσης των κηλίδων και της τοποθέτηση του πλέγματος..... | 89 |
| Σχήμα 4.3: Κατακόρυφη προβολή μια εικόνας μικροσυστοιχιών η οποία περιέχει τέσσερις συστάδες σε μια γραμμή, η μια δίπλα στην άλλη..... | 90 |
| Σχήμα 4.4: Εικόνα μικροσυστοιχιών η οποία περιέχει γραμμές και στήλες από συστάδες..... | 91 |
| Σχήμα 4.5: Τα αποτελέσματα των τεσσάρων βημάτων της επεξεργασίας των προβολών της εικόνας..... | 92 |
| Σχήμα 4.6: Η διαδικασία της επαλήθευσης των κοιλάδων που έχουν εντοπιστεί..... | 94 |
| Σχήμα 4.7: Οι προβολές μια εικόνας τετραγωνικού πλέγματος και μια εικόνας εξαγωνικού πλέγματος..... | 95 |
| Σχήμα 4.8: Καθώς το νερό γεμίζει την προβολή της εικόνας η επιφάνεια του λειτουργεί ως μεταβλητό κατώφλι. Τελικά το νερό σταματάει όταν μια κοιλάδα ξεχειλίζει από νερό (στο σημείο που επισημαίνεται με τον κύκλο. Στην προβολή δημιουργούνται μικρές και μεγάλες κοιλάδες..... | 96 |
| Σχήμα 4.9: Μια εικόνα εξαγωνικού πλέγματος στραμμένη στις 45° και στις 60° , και οι προβολές της σε αυτές τις διευθύνσεις..... | 97 |
| Σχήμα 4.10: Διάγραμμα ροής για τον εντοπισμό των υβριδοποιημένων κηλίδων..... | 99 |
| Σχήμα 4.11: Δυσδιάστατο πρότυπο οι φωτεινότητες του οποίου ακολουθούν κανονική κατανομή..... | 100 |
| Σχήμα 4.12: Συνδετικότητα 8-γειτονιάς..... | 103 |
| Σχήμα 4.13: Η διαδικασία εντοπισμού μη υβριδοποιημένων κηλίδων για εικόνες τετραγωνικού πλέγματος..... | 107 |
| Σχήμα 4.14: Οι τρεις ιδιότητες του πλέγματος που χρησιμοποιήσαμε για την ανάπτυξη του αλγορίθμου των αναπτυσσόμενων ομόκεντρων πολυγώνων. α) οποιαδήποτε κηλίδα είναι κέντρο πολυγώνου, β) οποιαδήποτε κηλίδα είναι κέντρο ομόκεντρων πολυγώνων γ) ο υπολογισμός των κηλίδων στο περίγραμμα το κάθε πολυγώνου μπορεί να υπολογιστεί..... | 109 |
| Σχήμα 4.15: Τα επίπεδα των ομόκεντρων πολυγώνων και ο αριθμός των κηλίδων επάνω στο περίγραμμά τους..... | 109 |

| | |
|--|-----|
| Σχήμα 4.16: Ψευδοκώδικας του αλγόριθμου των αναπτυσσόμενων ομόκεντρων πολυγώνων..... | 110 |
| Σχήμα 4.17: Η σχέση της επικρατούσας τιμής των αποστάσεων δύο γειτονικών κηλίδων με τις ακτίνες των ομόκεντρων εξαγώνων. | 111 |
| Σχήμα 4.18: Η διόρθωση των λαθών στρογγυλοποίησης | 115 |
| Σχήμα 4.19: Βελτιστοποίηση της θέσης του σημείου εκκίνησης του αλγορίθμου. | 117 |
| Σχήμα 4.20: Τέσσερα σημεία εκκίνησης του αλγορίθμου. | 118 |
| Σχήμα 4.21: Απλή απόδοση τετράφωνου κελιού σε κάθε εντοπισμένη κηλίδα. | 119 |
| Σχήμα 4.22: Voronoi διάγραμμα τυχαίων σημείων στο επίπεδο. | 120 |
| Σχήμα 4.23: Τα κέντρα των γειτονικών κηλίδων συνεισφέρουν σημαντικά ώστε το κελί να τοποθετηθεί σωστά..... | 121 |
| Σχήμα 4.24: Η τοποθέτηση τεχνιτών κέντρων γύρω από την συστάδα για τον περιορισμό του Βορονόι διαγράμματος..... | 122 |
| Σχήμα 4.25: Ιστογράμματα των οριζόντιων και κατακόρυφων αποστάσεων των κέντρων των εντοπισμένων με τα κέντρα της επισημείωσης. | 123 |
| Σχήμα 4.26: Ιστόγραμμα των ευκλείδειων αποστάσεων των κέντρων που εντοπίστηκαν με τα επισημειωμένα κέντρα. | 124 |
| Σχήμα 4.27: Τα αποτελέσματα της μεθοδολογίας για εικόνες με εξαγωνικό πλέγμα: α) μέρος της αρχικής εικόνας, β) το αποτέλεσμα του εντοπισμού των φωτεινών κηλίδων γ) το αποτέλεσμα του αλγόριθμου των αναπτυσσόμενων ομόκεντρων πολυγώνων δ) το αποτέλεσμα της εφαρμογής του Βορονόι διαγράμματος. | 126 |
| Σχήμα 4.28: Το ιστόγραμμα των αποστάσεων μεταξύ των εντοπισμένων κέντρων και των κέντρων της επισημείωσης. | 127 |
| Σχήμα 4.29: Το πλήθος των κέντρων για κάθε απόσταση μεταξύ των εντοπισμένων και επισημειωμένων κέντρων σε συνάρτηση με το πλήθος των σημείων εκκίνησης. | 130 |
| Σχήμα 4.30: Η ακρίβεια της μεθόδου, η μέση τιμή και η τυπική απόκλιση των αποστάσεων σε συνάρτηση με το πλήθος των σημείων εκκίνησης. | 131 |
| Σχήμα 4.31: Ιστογράμματα των αποστάσεων για κάθε ένα σύνολο εικόνων. | 133 |

| | |
|---|-----|
| Σχήμα 5.1: Διάγραμμα ροής της μεθοδολογίας για την κατάτμηση..... | 140 |
| Σχήμα 5.2: Η διαδικασία εξαγωγής των χαρακτηριστικών. Τα χαρακτηριστικά φωτεινότητας σχετίζονται με την φωτεινότητα του P_i και της 3x3 γειτονίας του και για τα δύο κανάλια της εικόνας. Τα χαρακτηριστικά θέσης είναι οι συντεταγμένες του εικονοστοιχείου P_i και οι αποστάσεις του εικονοστοιχείου από το κέντρο C. Τα χαρακτηριστικά σχήματος είναι ο συντελεστής συσχέτισης της 11X11 δισδιάστατης κανονικής κατανομής με την 11X11 γειτονιά του P_i και για τα δύο κανάλια της εικόνας. | 143 |
| Σχήμα 5.3: Δισδιάστατο παράδειγμα κατά το οποίο τα οι άξονες μετασχηματίζεται ελαχιστοποιώντας την διακύμανση..... | 146 |
| Σχήμα 5.4: Τα βήματα που εκτελεί επαναληπτικά ο αλγόριθμος K-μέσων | 152 |
| Σχήμα 5.5: Αποτέλεσμα της κατάτμησης που φανερώνει την αδυναμία του κανόνα υψηλότερης φωτεινότητας να αντιστοιχίσει σωστά τις παραγόμενες ομάδες με τις κατηγορίες εικονοστοιχείων: α) αποτέλεσμα σωστής αντιστοίχισης, β) αποτέλεσμα λανθασμένης αντιστοίχισης. | 154 |
| Σχήμα 5.6: Εύρεση του βέλτιστου υπερεπίπεδου για των διαχωρισμό των προτύπων από τις μηχανές διανυσμάτων υποστήριξης. Αριστερά φαίνονται τρία υπερεπίπεδα που χωρίζουν μπορούν να χωρίσουν τα πρότυπα, ενώ δεξιά φαίνεται το υπερεπίπεδο το οποίο μεγιστοποιεί το περιθώριο..... | 158 |
| Σχήμα 5.7: Παράδειγμα μη γραμμικών μηχανών διανυσμάτων υποστήριξης στον τρισδιάστατο χώρο. | 160 |
| Σχήμα 5.8: Η τεχνική δικτυωτής αναζήτησης για την βελτιστοποίησης δύο παραμέτρων. Η διαβαθμίσεις στο χρώμα του διαγράμματος αντιστοιχούν στην ακρίβεια της μεθόδου. Όσο πιο σκούρο είναι το χρώμα τόσο η μέθοδος παρουσιάζει μεγαλύτερη αξιοπιστία. Το ζεύγος που θα επιλέξει η μέθοδος για τις παραμέτρους θα είναι το ζεύγος το οποίο βρίσκεται στο πιο σκούρο σημείο του διαγράμματος. | 161 |
| Σχήμα 5.9: Τα αποτελέσματα της κατάτμησης των τριών υλοποιημένων μεθόδων για τρεις ενδεικτικές κηλίδες. | 164 |
| Σχήμα 5.10: Τα διαγράμματα διασποράς και τα διαγράμματα M-A των τριών μεθόδων βασισμένων σε τεχνικές ομαδοποίησης α) Διάγραμμα διασποράς του K-μέσων με 2 | |

χαρακτηρίστηκα και $K=2$, β) Διάγραμμα M-A του K-μέσων με 2 χαρακτηρίστηκα και $K=2$ γ) Διάγραμμα διασποράς του K-μέσων με 7 χαρακτηρίστηκα και $K=2$, δ) Διάγραμμα M-A του K-μέσων με 7 χαρακτηρίστηκα και $K=2$, ε) Διάγραμμα διασποράς του ασαφούς C-μέσων με 11 χαρακτηρίστηκα και $C=3$, στ) Διάγραμμα M-A του ασαφούς C-μέσων με 11 χαρακτηρίστηκα και $C=3$ 166

Σχήμα 5.11: Αριθμός σωστά κατατιμημένες κηλίδες για κάθε τιμή των τετραγωνικών σφαλμάτων του λόγου RAT. 169

Σχήμα 5.12: Οπτικοποίηση των αποτελεσμάτων της κατάτιμησης με επίβλεψη. 171

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

| | |
|--|-----|
| Πίνακας 1.1: Ο γενετικός κώδικας..... | 22 |
| Πίνακας 2.1: Διαφορετικές φθορίζουσες ουσίες που χρησιμοποιούνται στην τεχνολογία μικροσυστοιχιών. Τα μέγιστα μήκη κύματος διέγερσης και εκπομπής φαίνονται στην δεύτερη και την Τρίτη στήλη του πίνακα αντίστοιχα. | 41 |
| Πίνακας 4.1: Η ακρίβεια των κέντρων που εντοπιστήκαν με στα όρια της θεωρητικής κηλίδας, η μέση τιμή και η τυπική απόκλιση των αποστάσεων των κέντρων που εντοπίστηκαν με τα κέντρα της επισημείωσης..... | 125 |
| Πίνακας 4.2: Τα αποτελέσματα της προτεινόμενης μεθόδου με χρήση 100 αρχικών σημείων. | 128 |
| Πίνακας 4.3: Τα αποτελέσματα της μεθόδου για διαφορετικό πλήθος σημείων εκκίνησης. | 129 |
| Πίνακας 4.4: Αριθμητικά αποτελέσματα τις γενικευμένης μεθοδολογίας σε τέσσερα διαφορετικά σύνολα εικόνων. | 133 |
| Πίνακας 4.5: Συγκριτικά αποτελέσματα των προσεγγίσεων που αναπτύχθηκαν στην παρούσα διατριβή για τον εντοπισμό των κηλίδων, με χρήση και των τεσσάρων συνόλων εικόνων..... | 138 |
| Πίνακας 5.1: Τα χαρακτηριστικά που εξάγονται για κάθε εικονοστοιχείο της εικόνας... .. | 144 |
| Πίνακας 5.2: Τα αποτελέσματα της ευαισθησίας, της εξειδίκευσης και της ακρίβειας για τις πραγματικές εικόνες με χρήση του Μπεϋζιανού ταξινομητή και των Μηχανών Διανυσμάτων Υποστήριξης..... | 172 |
| Σχήμα 5.13: Οπτικοποίηση των αποτελεσμάτων της κατάτμησης με επίβλεψη. | 173 |
| Πίνακας 5.3: Ακρίβεια της μεθόδου για το πρόβλημα τριών κλάσεων των προσομοιωμένων εικόνων..... | 175 |
| Πίνακας 5.4: Συγκριτικά αποτελέσματα μεθόδων χωρίς επίβλεψη και μεθόδων με επίβλεψη..... | 176 |

ΕΙΣΑΓΩΓΗ

Το πρώτο κεφάλαιο της διατριβής επιχειρεί να παρουσιάσει όλες τις πτυχές της βιολογίας που απαιτούνται για την πλήρη κατανόηση του πειράματος των μικροσυστοιχιών. Συγκεκριμένα στο κεφάλαιο αυτό επιχειρείται μια συνοπτική περιγραφή των λειτουργιών του κυττάρου, αλλά και των ιδιοτήτων του DNA. Αυτό γίνεται με έμφαση στην ιδιότητα της υβριδοποίησης, στην οποία στηρίζεται η παρούσα τεχνολογία. Τέλος, το πρώτο κεφάλαιο αναφέρεται και σε διάφορες βιολογικές έννοιες, οι οποίες χρησιμοποιούνται στο φάσμα των εφαρμογών των μικροσυστοιχιών, όπως για παράδειγμα οι πολυμορφισμοί του DNA.

Στο δεύτερο κεφάλαιο περιγράφονται οι βασικές αρχές του πειράματος των μικροσυστοιχιών. Σκοπός του κεφαλαίου αυτού είναι η περιγραφή όλων των βημάτων και των διαδικασιών που προηγούνται της εξαγωγής και της επεξεργασίας των εικόνων. Τα βήματα αυτά αφορούν την κατασκευή των ίδιων των μικροσυστοιχιών αλλά και την διαδικασία του βιολογικού πειράματος. Τέλος κάνουμε μια συνοπτική περιγραφή των συνόλων δεδομένων, τα οποία χρησιμοποιήθηκαν κατά την διάρκεια της διατριβής με σκοπό την αξιολόγηση των μεθόδων οι οποίες αναπτύχθηκαν.

Το τρίτο κεφάλαιο παρουσιάζει την βιβλιογραφική ανασκόπηση (state-of-the-art) στο πεδίο της επεξεργασίας των εικόνων. Τρία είναι τα βασικά στάδια τα οποία περιλαμβάνει η επεξεργασία εικόνας, και αυτά περιγράφονται ξεχωριστά. Τα τρία αυτά στάδια είναι ο εντοπισμός των κηλίδων με παράλληλη τοποθέτηση ενός πλέγματος στην εικόνα, η κατάτμηση των κηλίδων, και τέλος η εξαγωγή των ποσοτικοποιημένων τιμών. Σκοπός του κεφαλαίου είναι να γίνει μια αναλυτική ανασκόπηση της βιβλιογραφίας, περιγράφοντας όλες τις εργασίες που έχουν αναπτυχθεί από τους ερευνητές και για τρία στάδια της επεξεργασίας των εικόνων. Ιδιαίτερη έμφαση δίνεται στην εξελικτική πορεία της έρευνας στο πεδίο αυτό.

Το τέταρτο κεφάλαιο περιγραφεί αναλυτικά την έρευνα που διεξήχθει στα πλαίσια της παρούσας διατριβής, για τον εντοπισμό των κηλίδων και την τοποθέτηση του πλέγματος στις εικόνες. Όπως θα δούμε κατά την διάρκεια της διατριβής ασχοληθήκαμε με εικόνες όπου προέρχονται από τις δύο κατασκευαστικές προσεγγίσεις των μικροσυστοιχιών. Αυτές οι προσεγγίσεις αφορούν την χωροθέτηση των βιολογικών αλληλουχιών επάνω στην μικροσυστοιχία, η οποία γίνεται είτε με τετραγωνική είτε με εξαγωνική δομή. Η παρούσα διατριβή καταλήγει σε μια ενιαία μεθοδολογία για τον εντοπισμό των κηλίδων σε εικόνες και από τις δύο προσεγγίσεις.

Στο πέμπτο κεφάλαιο αναπτύσσεται η προτεινόμενη μεθοδολογία για την κατάτμηση των εικόνων. Για τον διαχωρισμό των εικονοστοιχείων της εικόνας σε εικονοστοιχεία σήματος ή εικονοστοιχεία υποβάθρου χρησιμοποιήθηκαν τεχνικές Μηχανικής Μάθησης (Machine Learning). Η παρούσα διατριβή έχει προσφέρει τρεις μεγάλες καινοτομίες στο στάδιο της κατάτμησης. Η πρώτη αφορά το διάνυσμα των χαρακτηριστικών που τροφοδοτεί τους αλγορίθμους, το οποίο διευρύνθηκε σημαντικά για την εισαγωγή χαρακτηριστικών περάν αυτών που εξάγονται από την φωτεινότητα του κάθε εικονοστοιχείου. Η δεύτερη καινοτομία είναι η εισαγωγή της τρίτης κλάσης εικονοστοιχείων. Εκτός από τις κλάσεις του σήματος και του υποβάθρου, εισάγαμε μια τρίτη κλάση, με σκοπό να συμπεριληφθούν σε αυτήν όλα εκείνα τα εικονοστοιχεία που δεν θέλουμε να λάβουν μέρος στον υπολογισμό των ποσοτικοποιημένων τιμών. Εκτός από τα παραπάνω, για πρώτη φορά έγινε χρήση τεχνικών ταξινόμησης για τον διαχωρισμό των εικονοστοιχείων, αυξάνοντας σε μεγάλο βαθμό την ακρίβεια της μεθοδολογίας.

Το τελευταίο κεφάλαιο συνοψίζει τα αποτελέσματα της έρευνας που διεξήγαμε κατά την διάρκεια της παρούσας διδακτορικής διατριβής, καταλήγοντας σε χρήσιμα συμπεράσματα για την συνεισφορά της στο συγκεκριμένο πεδίο έρευνας.

Κεφάλαιο 1^ο: Βιολογικό Υπόβαθρο

1.1. Βασικές Έννοιες

Βασική προϋπόθεση για την κατανόηση όλων των διαδικασιών που διενεργούνται σε ένα πείραμα μικροσυστοιχιών είναι η γνώση των βασικών αρχών της βιολογίας που διέπουν αυτήν την τεχνολογία. Όσον αφορά στις βιολογικές διαδικασίες είναι πολλά αυτά που πρέπει να αποσαφηνιστούν, ξεκινώντας από τον ρόλο των γονιδίων και του DNA στους ζώντες οργανισμούς, και συνεχίζοντας με τις διαδικασίες της αντιγραφής του, της έκφρασης των γονιδίων αλλά και της υβριδοποίησης των συμπληρωματικών αλυσίδων.

Η ικανότητα των κυττάρων να αποθηκεύουν, να ανακαλούν και να μεταφράζουν τις γενετικές πληροφορίες συνιστά ίσως την πιο σημαντική ιδιότητα για την δημιουργία και την συντήρηση ενός ζωντανού οργανισμού. Κατά την κυτταρική διαίρεση, όλες οι κληρονομικές πληροφορίες μεταβιβάζονται από ένα κύτταρο στα θυγατρικά του κύτταρα, και κατ' αυτόν τον τρόπο από έναν οργανισμό στους απογόνους του μέσω των αναπαραγωγικών κυττάρων [1-4]. Κάθε κύτταρο αποθηκεύει τις γενετικές πληροφορίες ως γονίδια (genes). Αυτά είναι και τα οποία καθορίζουν τα χαρακτηριστικά ενός είδους στο σύνολο του αλλά και τα χαρακτηριστικά των επιμέρους οργανισμών που ανήκουν στο είδος αυτό.

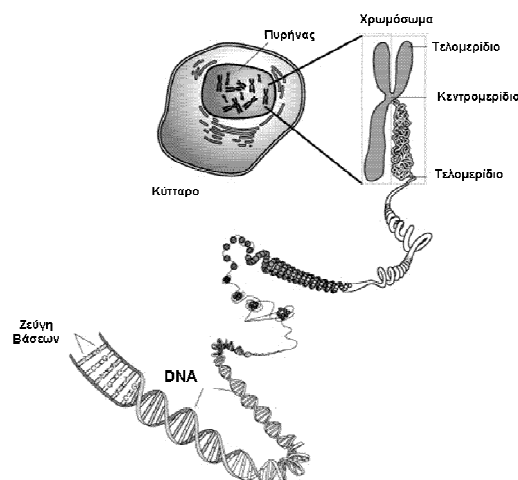
Η σύσταση των γονιδίων αποτέλεσε στις αρχές του εικοστού αιώνα ένα πεδίο έρευνας που απασχόλησε σε μεγάλο βαθμό τους ερευνητές. Διάφορες ιδιότητες ήλθαν το επίκεντρο της έρευνας, όπως για παράδειγμα η αντιγραφή των γενετικών πληροφοριών σε εκατομμύρια αντίγραφα, η ακρίβεια στην αντιγραφή, αλλά και ο τρόπος με τον οποίο αποθηκεύονται όλες αυτές οι πληροφορίες στον περιορισμένο χώρο ενός κυττάρου. Τα αποτελέσματα των πρώτων μελετών άρχισαν να διαφαίνονται στα μέσα περίπου του 20^{ου} αιώνα, όπου από έρευνες που έγιναν κυρίως σε μύκητες διαπιστώθηκε ότι οι γενετικές αυτές πληροφορίες αποτελούνται κυρίως από οδηγίες για την δημιουργία πρωτεϊνών. Οι πρωτεΐνες [5-6] είναι μακρομόρια, τα οποία επιτελούν τις περισσότερες από τις κυτταρικές λειτουργίες. Πιο συγκεκριμένα, οι πρωτεΐνες λειτουργούν ως δομικοί λίθοι για διάφορες κυτταρικές δομές, δρουν ως ένζυμα που καταλύουν όλες τις χημικές αντιδράσεις του κυττάρου, ρυθμίζουν την έκφραση των γονιδίων, επιτρέπουν στα κύτταρα την μετακίνηση αλλά και την επικοινωνία μεταξύ τους. Οι πρωτεΐνες που μπορεί να συνθέσει ένα κύτταρο είναι αυτές που καθορίζουν τις ιδιότητες του, αλλά και τις λειτουργίες του. Η άλλη καθοριστική ανακάλυψη που σημειώθηκε εκείνη την περίοδο ήταν η ταυτοποίηση του δεοξυριβονουκλεϊκού οξέος (DeoxyriboNucleic Acid – DNA) [7-8] ως πιθανού φορέα των γενετικών πληροφοριών. Ο μηχανισμός όμως κατά τον οποίο αντιγράφονται οι γενετικές πληροφορίες για να μεταφερθούν από ένα κύτταρο σε ένα άλλο, αλλά και ο ρόλος των οδηγιών που περιέχονται στο DNA για την σύνθεση των πρωτεϊνών παρέμεναν άγνωστα μέχρι που οι Watson και Crick το 1953 [9] καθόρισαν την δομή του DNA. Η δομή του DNA έλυσε αμέσως το ανεξήγητο του τρόπου αντιγραφής του και πρόσφερε τις πρώτες ενδείξεις για το πώς ένα μόριο DNA μπορεί να κωδικοποιεί τις οδηγίες για την σύνθεση των πρωτεϊνών. Στα παρακάτω εδάφια, θα εξετάσουμε τον τρόπο με τον οποίο η δομή του DNA επιτρέπει την αξιόπιστη αντιγραφή και διατήρηση των γενετικών οδηγιών, αλλά και τον τρόπο με τον οποίο ένα μόριο DNA μπορεί να κωδικοποιεί τις πληροφορίες για την σύνθεση και εξειδίκευση των πρωτεϊνών.

Τα κύτταρα δεν περιέχουν μόνο έναν περίτεχνο εξοπλισμό για την ακριβή αντιγραφή των γενετικών πληροφοριών αλλά και εξειδικευμένα ένζυμα για την επιδιόρθωση των λαθών που μπορούν συμβούν κατά την αντιγραφή του DNA. Παρά την ασφαλιστική αυτή δικλείδα, αυτών των συστημάτων που προστατεύουν τις γενετικές πληροφορίες από λάθη

κατά την αντιγραφή και από τυχαίες βλάβες, ορισμένες φορές συμβαίνουν μόνιμες αλλαγές ή μεταλλάξεις (mutations). Καμιά φορά η μεταλλάξεις του DNA, οι οποίες συχνά επηρεάζουν τις πληροφορίες που κωδικοποιεί, μπορεί να ωφελήσει έναν οργανισμό. Κάποιες μεταλλάξεις μπορεί περιστασιακά για παράδειγμα να καταστήσουν τα βακτήρια περισσότερο ανθεκτικά στα αντιβιοτικά που χρησιμοποιούνται για την καταπολέμηση τους. Επίσης, η συσσώρευση μεταλλάξεων στο DNA που συμβαίνουν σε μεγάλης διάρκειας χρονικά διαστήματα είναι αυτή που συμβάλει στην ποικιλότητα που κάνει το ένα είδος διαφορετικό από το άλλο, αλλά και στην διαφορετικότητα των οργανισμών που ανήκουν στο ίδιο είδος. Ωστόσο, οι περισσότερες μεταλλάξεις είναι επιβλαβείς για έναν οργανισμό εφόσον συνδέονται με την εμφάνιση των κληρονομικών νοσημάτων. Μεταλλάξεις που συμβαίνουν στα κύτταρα του σώματος κατά την διάρκεια της ζωής ενός ανθρώπου, μπορούν επίσης να προκαλέσουν κάποια επίκτητη νόσο, με χαρακτηριστικότερο παράδειγμα να αποτελούν οι διάφορες μορφές του καρκίνου [10-13].

1.2. Η δομή και η λειτουργία του DNA

Όπως αναφέρθηκε ήδη οι βιολόγοι είχαν αναγνωρίσει την ύπαρξη των γονιδίων στις αρχές περίπου του 20^{ου} αιώνα μέσα στα χρωμοσώματα, τα οποία ήταν τότε γνωστά ως νηματοειδείς δομές στον πυρήνα των ευκαρυωτικών κυττάρων, όπως φαίνεται και στο Σχήμα 1.1.

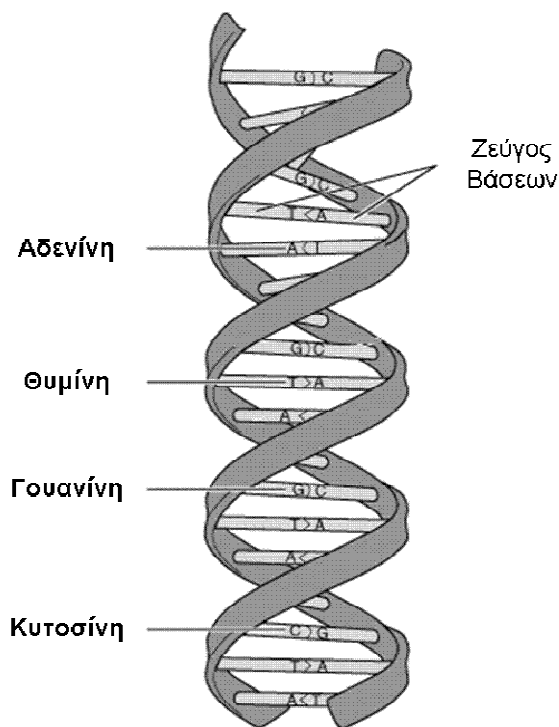


Σχήμα 1.1: Η ύπαρξη των γονιδίων στα χρωμοσώματα των ζωντανών οργανισμών.

Όταν ξεκίνησαν οι βιοχημικές αναλύσεις των χρωμοσωμάτων βρέθηκε ότι αυτά αποτελούνται από DNA και από πρωτεΐνες. Δεδομένου ότι το μόριο του DNA φάνταζε ένα πολύ απλό μόριο, επικράτησε η άποψη ότι τα γονίδια έπρεπε να αποτελούνται από πρωτεΐνες, οι οποίες έχουν πολύ μεγαλύτερη χημική ποικιλομορφία. Ωστόσο με την πάροδο του χρόνου εξετάζοντας την δομή και τις χημικές ιδιότητες του DNA αποδείχθηκε ότι παρά τη χημική απλότητα του, είναι ιδανικό για να είναι αυτό η πρώτη ύλη για τα γονίδια. Έτσι, από πειράματα που έγιναν σε μια μεγάλη ποικιλία οργανισμών, φτάσαμε στο συμπέρασμα ότι τα γονίδια κάθε κυττάρου αποτελούνται από DNA.

Στις μέρες μας, το γεγονός ότι το DNA είναι το γενετικό υλικό είναι τόσο θεμελιώδες για τη βιολογική σκέψη, ώστε να φαίνεται δύσκολο να συνειδητοποιήσουμε το τεράστιο νοητικό κενό που κάλυψε αυτή η ανακάλυψη. Ωστόσο, στα μέσα του προηγούμενου αιώνα οι βιολόγοι δεν μπορούσαν να αποδεχτούν ότι το γενετικό υλικό είναι το DNA, και αυτό λόγω της χημικής του σύστασης η οποία κρίθηκε εξαιρετικά απλή, για ένα μόριο το οποίο έχει την μεγαλύτερη ευθύνη για την δημιουργία της ζωής. Και είναι γεγονός ότι το DNA είναι ένα μακρύ πολυμερές, αποτελούμενο από τέσσερα μόνο είδη δομικών μονάδων, οι οποίες μάλιστα μοιάζουν μεταξύ τους από χημικής άποψης. Όσο η τεχνολογία εξελισσόταν το DNA υποβλήθηκε σε διάφορα πειράματα, ένα εκ των οποίων ήταν ο καθορισμός της τρισδιάστατης ατομικής δομής του μορίου του με ανάλυση της περίθλασης ακτίνων X [14]. Το πείραμα αυτό έδειξε ότι το DNA αποτελείται από δύο κλώνους (double-stranded) που περιελίσσονται σε μια έλικα, συμπέρασμα το οποίο ήταν αποφασιστικής σημασίας. Πάνω σε αυτό το συμπέρασμα στηρίχθηκε και το μοντέλο που πρότειναν οι Watson και Crick για την δομή του DNA, και το οποίο κατέστησε αντιληπτές τις δυνατότητες του DNA για την αντιγραφή του και την αποθήκευση της γενετικής πληροφορίας. Η δίκλωνη αλυσίδα του DNA παρουσιάζεται στο Σχήμα 1.2.

Όπως φαίνεται και στο σχήμα 1.2 ένα μόριο DNA αποτελείται από δυο μακριές πολυνουκλεοτιδικές αλυσίδες, οι οποίες με την σειρά τους αποτελούνται από τέσσερα είδη νουκλεοτιδίων. Οι αλυσίδες αυτές αναφέρονται ως αλυσίδες DNA (DNA chains) ή κλώνοι DNA (DNA strands) και συνδέονται με δεσμούς υδρογόνου μεταξύ των βάσεων των νουκλεοτιδίων. Τα νουκλεοτίδια αποτελούνται από μία πεντόζη, με την οποία συνδέονται μία ή περισσότερες φωσφορικές ομάδες, και μία αζωτούχα βάση.

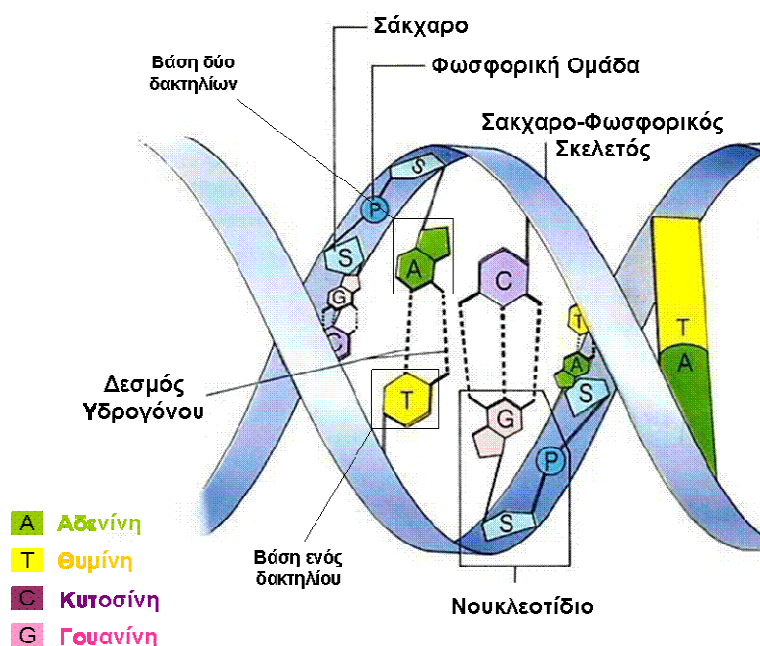


Σχήμα 1.2: Η δίκλωνη αλυσίδα του DNA.

Στην περίπτωση των νουκλεοτιδίων του DNA η πεντόζη που συνδέεται με μία φωσφορική ομάδα είναι η δεοξυριβόζη, ενώ η βάση μπορεί να είναι η αδενίνη (A), η κυτοσίνη (C), η γουανίνη (G) ή η θυμίνη (T). Τα νουκλεοτίδια συνδέονται ομοιοπολικά μεταξύ τους σε μία αλυσίδα μέσω των σακχάρων και των φωσφορικών ομάδων τους, δημιουργώντας έτσι ένα σκελετό από εναλλασσόμενες μονάδες σακχάρου-φωσφορικής ομάδας-σακχάρου-φωσφορικής ομάδας, ο οποίος ονομάζεται σάκχαρο-φωσφορικός σκελετός (Sugar-Phosphate backbone). Οι τέσσερις δομικές μονάδες διαφέρουν μεταξύ τους μόνο ως προς τη βάση τους. Έτσι τα ίδια σύμβολα που χρησιμοποιούνται για τις βάσεις (A, C, G και T) χρησιμοποιούνται επίσης και για τα τέσσερα διαφορετικά νουκλεοτίδια, δηλαδή τις βάσεις μαζί με τα σάκχαρα και τις φωσφορικές ομάδες τους.

Οι δύο αλυσίδες στη διπλή έλικα (double helix) του DNA συνδέονται με δεσμούς υδρογόνου μεταξύ των βάσεων των διαφορετικών κλώνων. Έτσι, όλες οι βάσεις είναι τοποθετημένες στο εσωτερικό της έλικας, ενώ τα σάκχαρα και η φωσφορική ομάδα στο εξωτερικό της, όπως ακριβώς φαίνεται στο σχήμα 1.3. Μόνο οι συμπληρωματικές βάσεις, δηλαδή η A με την T και η G με την C, συνδέονται μεταξύ τους με δεσμούς υδρογόνου.

Αυτό ουσιαστικά συμβαίνει λόγω της σύνδεσης μίας ογκώδους βάσης με δύο δακτυλίους (μία βάση πουρίνης) με μία βάση ενός δακτυλίου (μία βάση πυριμιδίνης).



Σχήμα 1.3: Η δομή του DNA

Ο τρόπος με τον οποίο συνδέονται μεταξύ τους τα νουκλεοτίδια, προσδίδει σε έναν κλώνο DNA χημική πολικότητα. Καθάρά για λόγους σύμβασης, και λόγω περαιτέρω λεπτομερειών στους χημικούς δεσμούς των νουκλεοτιδικών μονάδων, έχει επικρατήσει τα δύο άκρα μιας αλυσίδας να συμβολίζονται ως 3' το ένα και ως 5' το άλλο. Τα ζεύγη των βάσεων τοποθετούνται στο εσωτερικό της διπλής έλικας, και εκεί όπου ελαχιστοποιείται η ενέργεια. Δεδομένου ότι κάθε ζεύγος βάσεων έχει παρόμοιο εύρος, διατηρεί τον σακχαροφωσφορικό σκελετό σε ίση απόσταση κατά μήκος του μορίου του DNA, ενώ οι σακχαροφωσφορικοί σκελετοί περιελίσσονται ο ένας γύρω από τον άλλο για να σχηματίσουν μία διπλή έλικα. Αυτό συμβαίνει μόνο όταν οι δύο κλώνοι της είναι αντιπαράλληλοι (antiparallel), δηλαδή μόνο όταν η πολικότητα του ενός κλώνου έχει αντίθετο προσανατολισμό από την πολικότητα του άλλου. Έτσι, ο κάθε κλώνος ενός μορίου DNA περιέχει μία αλληλουχία νουκλεοτιδίων, η οποία είναι ακριβώς συμπληρωματική (complementary) προς την αλληλουχία των νουκλεοτιδίων του άλλου κλώνου, γεγονός το οποίο διαδραματίζει ουσιαστικό ρόλο στην αντιγραφή του DNA. Στην συμπληρωματικότητα των δύο κλώνων βασίζεται και η ιδιότητα της υβριδοποίησης, η οποία είναι και ακρογωνιαίος λίθος της τεχνολογίας των μικροσυστοιχιών.

Ένα πολύ σημαντικό ζήτημα αφορά στην μεταφορά και την κληρονομικότητα γενετικών πληροφοριών με την χημική τους μορφή για την δημιουργία ενός οργανισμού. Την στιγμή δηλαδή που ένα κύτταρο διαιρείται, οι βιολογικές πληροφορίες που περιέχονται στα γονίδια πρέπει να αντιγραφούν με ακρίβεια και να μεταβιβαστούν στα δύο θυγατρικά κύτταρα. Όπως αναφέρθηκε και παραπάνω η εξήγηση δίδεται από την διαλεύκανση της δομής της διπλής έλικας του DNA.

Η αλληλουχία των νουκλεοτιδίων στο DNA είναι αυτή που καθορίζει την κωδικοποίηση των πληροφοριών κατά μήκος μιας αλυσίδας. Μπορούμε να υποθέσουμε ότι υπάρχει ένα αλφάβητο, το οποίο αποτελείται από τέσσερα μόνο γράμματα (τα γράμματα της κάθε βάσης A, C, G, T) και με το οποίο καταγράφονται τα βιολογικά μηνύματα στη χημική δομή του DNA. Οι οργανισμοί διαφέρουν μεταξύ τους επειδή τα αντίστοιχα μόρια του DNA τους έχουν διαφορετική αλληλουχία νουκλεοτιδίων και, επομένως, περιέχουν διαφορετικά βιολογικά μηνύματα.

Τα γονίδια επίσης περιέχουν τις οδηγίες για την παραγωγή των πρωτεϊνών, τον τρόπο δηλαδή με τον οποίο αυτές κωδικοποιούνται. Οι ιδιότητες μιας πρωτεΐνης, οι οποίες είναι υπεύθυνες για τη βιολογική λειτουργία της καθορίζονται από την τρισδιάστατη δομή της ,ενώ η δομή της πρωτεΐνης με τη σειρά της, καθορίζεται από την αλληλουχία των αμινοξέων από τα οποία αποτελείται. Έτσι, η γραμμική αλληλουχία των νουκλεοτιδίων ενός γονιδίου πρέπει με κάποιο τρόπο να προδιαγράφει τη γραμμική αλληλουχία των αμινοξέων μιας πρωτεΐνης. Ωστόσο, η ακριβής αντιστοιχία ανάμεσα στο αλφάβητο των τεσσάρων νουκλεοτιδίων του DNA και στο αλφάβητο των είκοσι αμινοξέων των πρωτεϊνών δεν είναι προφανής από τη δομή του DNA. Τα περισσότερα γονίδια είναι βραχέα τμήματα DNA που κωδικοποιούν, κατά περίπτωση, μια μόνο πρωτεΐνη. Ωστόσο, το τμήμα του DNA ενός γονιδίου δεν χρησιμοποιείται στο σύνολο του για την κωδικοποίηση της πρωτεΐνης που καθορίζεται από το γονίδιο αυτό. Ένα μεγάλο τμήμα του είναι επιφορτισμένο να καθορίζει πότε και σε ποιες ποσότητες θα παράγεται η πρωτεΐνη που κωδικοποιείται από το γονίδιο, οι οποίες ονομάζονται ρυθμιστικές περιοχές του γονιδίου.

Όλες οι πληροφορίες που περιέχει το DNA ενός οργανισμού ονομάζονται γονιδίωμα (genome). Γονιδίωμα επίσης ονομάζεται αντίστοιχα και το DNA που περιέχει αυτές τις πληροφορίες. Μία τάξη μεγέθους των πληροφοριών που περιέχονται στο γονιδίωμα μπορεί να δοθεί από τον αριθμό των νουκλεοτιδίων του. Το ανθρώπινο γονιδίωμα για παράδειγμα περιέχει περίπου τρία δισεκατομμύρια (3×10^9) νουκλεοτίδια, ενώ αριθμός των νουκλεοτιδίων ενός μικρού γονιδίου του ανθρώπου είναι της τάξης των μερικών εκατοντάδων.

1.3. Αντιγραφή του DNA

Κατά την διάρκεια της διαίρεσης του κυττάρου, το DNA του αντιγράφεται για να μεταδώσει τις γενετικές πληροφορίες στον νέο κύτταρο. Όπως αναφέρθηκε και ανωτέρω η δομή της διπλής έλικας είναι αυτή που εξηγεί τον τρόπο με τον οποίο το DNA παράγει ακριβή αντίγραφα.

Οι δύο κλώνοι του DNA αποτελούνται από αλληλουχία συμπληρωματικών βάσεων. Έτσι κάθε ένας από τους κλώνους μπορεί να λειτουργήσει ως εκμαγείο (template) για να δημιουργήσει μία νέα συμπληρωματική αλυσίδα. Αρχικά λοιπόν οι δύο κλώνοι του DNA διαχωρίζονται και ο καθένας τους ξεχωριστά δημιουργεί τον συμπληρωματικό του, αντιγράφοντας έτσι ακριβώς όλες τις γενετικές πληροφορίες με μια πολύ απλή διαδικασία. Εφόσον κάθε αρχικός κλώνος λειτουργεί ως εκμαγείο για τη σύνθεση ενός νέου κλώνου, η καθεμία από τις δυο νέες διπλές έλικες του DNA θα αποτελείται από έναν αρχικό κλώνο και από έναν άλλο, εντελώς νέο κλώνο. Μεγάλη εντύπωση βέβαια προκαλείται και από την ποσότητα των νουκλεοτιδίων που πρέπει να αντιγραφούν σε ένα μικρό διάστημα ώστε να λάβει χώρα η αντιγραφή ολόκληρου του γονιδιώματος. Το επίτευγμα αυτό πραγματοποιείται από ένα σύμπλεγμα πρωτεϊνών, οι οποίες, σχηματίζουν ένα είδος «αντιγραφικής μηχανής» (“replication machine”). Με την αντιγραφή του DNA από το αρχικό μόριο του DNA παράγονται δυο πλήρεις διπλές έλικες, στην οποία με εξαίρεση λίγων λαθών, η καθεμία έχει την ίδια αλληλουχία νουκλεοτιδίων με την αρχική.

Προκειμένου να χρησιμοποιηθούν ως εκμαγεία, οι δύο κλώνοι απαιτείται πρώτα η διπλή έλικα να διαχωριστεί ώστε να εκτεθούν αζευγάρωτες οι βάσεις. Οι δύο κλώνοι της διπλής έλικας του DNA συνδέονται με δεσμούς υδρογόνου, οι οποίοι μεν μπορεί να είναι ασθενείς μεμονωμένα, ωστόσο δεδομένου ότι είναι πάρα πολύ σε αριθμό κρατούν τους δύο κλώνους σταθερά συνδεδεμένους. Για τον διαχωρισμό των δύο κλώνων απαιτείται θερμοκρασία παραπλήσια με το σημείο ζέσεως του νερού, ώστε να απελευθερωθεί η απαιτούμενη ενέργεια για την διάσπαση των δεσμών υδρογόνου. Επομένως, κάτι άλλο είναι αυτό που πρέπει να δώσει την ώθηση ώστε να ξεκινήσει η διαδικασία της αντιγραφής. Τον ρόλο αυτό διαδραματίζουν ειδικές εναρκτήριες πρωτεΐνες (initiator proteins), οι οποίες σκοπεύουν στην σύνδεση τους με το DNA για να διασπάσουν τους δεσμούς υδρογόνου ανάμεσα στις βάσεις.

Ο διαχωρισμός της διπλής έλικας δεν συμβαίνει ταυτόχρονα αλλά ξεκινά από συγκεκριμένες θέσεις του DNA, οι οποίες καλούνται αφετηρίες αντιγραφής (replication origins) και χαρακτηρίζονται από μια συγκεκριμένη αλληλουχία νουκλεοτιδίων. Αυτές αποτελούνται από αλληλουχίες DNA που προσελκύουν τις εναρκτήριες πρωτεΐνες όπως επίσης και από τμήματα DNA που διανοίγονται πολύ εύκολα. Είναι γεγονός ότι ένα ζεύγος βάσεων A-T συνδέεται με λιγότερους δεσμούς υδρογόνου από ένα ζεύγος G-C (βλ. σχήμα 1.3), και άρα ένα τμήμα DNA πλούσιο σε ζεύγη Βάσεων A-T ανοίγει σχετικά εύκολα. Με αυτό το δεδομένο οι αφετηρίες αντιγραφής συνήθως είναι αλληλουχίες DNA πλούσιες σε A-T.

Δεδομένου ότι το DNA ξετυλίγεται ξεκινώντας από τις αφετηρίες και συνεχίζοντας στις υπόλοιπες βάσεις, μπορεί κανείς να υποθέσει ότι η απλοί οργανισμοί, το γονιδίωμα των οποίων αποτελείται από εκατομμύρια (και όχι δισεκατομμύρια βάσεις όπως το ανθρώπινο γονιδίωμα) μπορούν να αντιγράψουν το DNA τους με πολλαπλάσια ταχύτητα απ' ότι οι πολυπλοκότεροι οργανισμοί. Ωστόσο, ενώ οι απλοί οργανισμοί έχουν μόνο μία αφετηρία αντιγραφής στο DNA τους, το ανθρώπινο DNA περιέχει περίπου 10,000 αφετηρίες στις οποίες ξεκινάει διαδικασία της αντιγραφής σχεδόν ταυτόχρονα. Κατ' αυτόν τον τρόπο η ταχύτητα της αντιγραφής αυξάνεται σημαντικά.

Όταν μια εναρκτήρια πρωτεΐνη συνδεθεί στην αφετηρία αντιγραφής του DNA και διαχωρίσει τοπικά τους δύο κλώνους, προσελκύει μια ομάδα πρωτεϊνών, καθεμία από τις οποίες διαπεραιώνει μια ειδική λειτουργία, με σκοπό να διενεργήσουν την αντιγραφή του DNA.

Στα σημεία ακριβώς των αφετηριών της αντιγραφής όπου οι δύο κλώνοι της έλικας αρχίζουν να διαχωρίζονται δημιουργούνται διχάλες αντιγραφής (replication forks). Από τις διχάλες αυτές οι πρωτεΐνες μετακινούνται κατά μήκος του DNA, και χρησιμοποιούν τον κάθε κλώνο ως εκμαγείο για τη σύνθεση ενός θυγατρικού κλώνου. Είναι εύκολο να κατανοήσει κανείς ότι σε κάθε αφετηρία αντιγραφής δημιουργούνται δυο διχάλες, που απομακρύνονται από την αφετηρία προς αντίθετες κατευθύνσεις διαχωρίζοντας την διπλή έλικα του DNA. Γι' αυτόν ακριβώς των λόγο η αντιγραφή ονομάζεται αντιγραφή διπλής κατεύθυνσης (bidirectional).

Το σύμπλεγμα πρωτεϊνών, το οποίο ανοικοδομεί την νέα αλληλουχία περιέχει ένα ένζυμο γνωστό ως DNA πολυμεράση (DNA polymerase) που συνθέτει το νέο DNA χρησιμοποιώντας ως εκμαγείο έναν από τους παλαιούς κλώνους. Το ένζυμο καταλύει την προσθήκη νουκλεοτιδίων στο 3' άκρο ενός επιμηκυνόμενου κλώνου DNA, μέσω του σχηματισμού ενός φωσφοδιεστερικού δεσμού μεταξύ του 3' άκρου και της 5'-φωσφορικής ομάδας του εισερχόμενου νουκλεοτιδίου. Η DNA πολυμεράση δεν αποσυνδέεται από το DNA κάθε φορά που προσθέτει ένα νουκλεοτίδιο στην αυξανόμενη αλυσίδα, αντίθετα, παραμένει συνδεδεμένη με το DNA και μετακινείται κατά μήκος του DNA βήμα-βήμα για πολλούς κύκλους της αντίδρασης πολυμερισμού.

Δύο βασικές ιδιότητες της DNA πολυμεράσης είναι καλό να αναφερθούν σε αυτό το σημείο. Η πρώτη έχει να κάνει με το γεγονός ότι η DNA πολυμεράση καταλύει την αύξηση μιας αλυσίδας DNA μόνο προς την μια κατεύθυνση. Αντίθετα, είδαμε ότι ο σακχαρο-φωσφορικός σκελετός κάθε κλώνου μιας διπλής έλικας DNA έχει μια συγκεκριμένη χημική κατεύθυνση ή πολικότητα, που καθορίζεται από τον τρόπο με τον οποίο το κάθε σάκχαρο συνδέεται με το επόμενο. Επίσης είδαμε ότι οι δύο κλώνοι της διπλής έλικας έχουν αντίθετη κατεύθυνση. Με αυτόν τον τρόπο ο ένας νέος κλώνος DNA

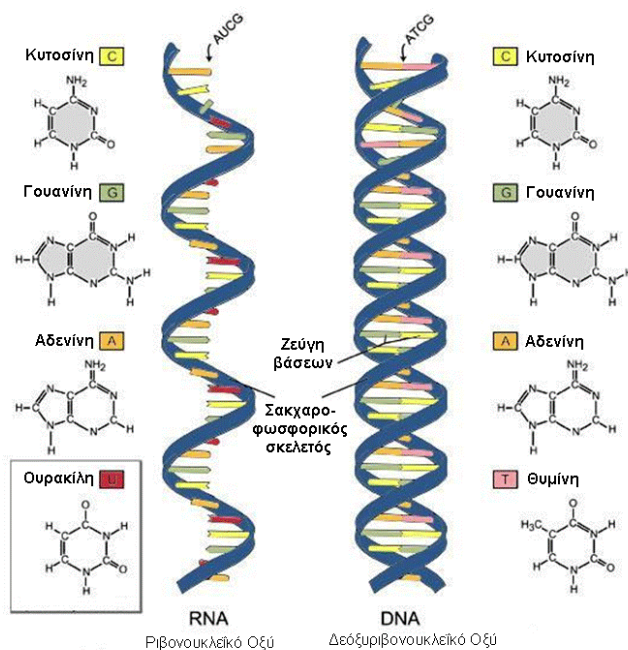
δημιουργείται πάνω σε εκμαγείο που έχει κατεύθυνση (3'-5'), ενώ ο άλλος κλώνος πάνω σε εκμαγείο που έχει την αντίθετη κατεύθυνση (5'-3').

Η DNA πολυμεράση προσθέτει νέες νουκλεοτιδικές μονάδες μόνο στο 3' άκρο της αλυσίδας. Θα περίμενε κανείς να υπάρχει και ένα δεύτερο ένζυμο το οποίο να δημιουργεί της νουκλεοτιδικές μονάδες στην αντίθετη κατεύθυνση. Δεδομένου όμως ότι τέτοιο ένζυμο δεν υπάρχει στην φύση ο κλώνος του DNA που πρέπει να συντίθεται ασυνεχώς με την αντίθετη κατεύθυνση, σε διαδοχικά μικρά τμήματα, με την DNA πολυμεράση να λειτουργεί σε αντίθετη κατεύθυνση από την κατεύθυνση της διχάλας αντιγραφής. Ο κλώνος του DNA που συντίθεται ασυνεχώς κατ' αυτόν τον τρόπο αποκαλείται καθυστερημένος κλώνος (lagging strand), ενώ ο κλώνος που συντίθεται συνεχώς αποκαλείται προπορευόμενος κλώνος (leading strand).

Η δεύτερη σημαντική ιδιότητα της DNA πολυμεράσης αναφέρεται στο γεγονός ότι μπορεί να διορθώνει τα λάθη που μπορεί να συμβούν κατά την διάρκεια της αντιγραφής. Παρά το γεγονός ότι η αντιγραφή λαμβάνει χώρα με πολύ μεγάλη ακρίβεια η πιστότητας αυξάνεται ακόμη περισσότερο από την ιδιότητα αυτή της DNA πολυμεράσης. Είναι ενδεικτικό να αναφέρουμε ότι κατά την αντιγραφή του DNA μπορεί να συμβεί ένα λάθος μόνο στα 10 εκατομμύρια νουκλεοτίδια.. Αν και όπως αναφέρουμε δύο συμπληρωματικές βάσεις συνδέονται μεταξύ τους, ωστόσο υπάρχουν και άλλα ζεύγη που μπορούν να συνδεθούν μεταξύ των τεσσάρων βάσεων, τα οποία όμως συνδέονται με πολύ ασθενέστερους δεσμούς. Τέτοια ζεύγη για παράδειγμα είναι τα ζεύγη G-T και C-A, τα οποία βέβαια συμβαίνουν πολύ σπάνια. Αν όμως τέτοια λάθη αρχίσουν να συσσωρεύονται με την πάροδο του χρόνου και μετά από πολλές αντιγραφές του DNA μπορεί να αποβούν μοιραία για την ζωή του κυττάρου. Η DNA πολυμεράση έχει την ιδιότητα να ελέγχει αν το νουκλεοτίδιο που έχει προστεθεί προηγουμένως ζευγαρώνει με το σωστό νουκλεοτίδιο του εκμαγείου. Μόνο αν το ζεύγος των βάσεων που έχει δημιουργηθεί είναι σωστό, η πολυμεράση θα προχωρήσει να προσθέτει τα επόμενα νουκλεοτίδια. Στην αντίθετη περίπτωση, το ένζυμο αφαιρεί το αταίριαστο νουκλεοτίδιο, διασπώντας τον φωσφοδιεστερικό δεσμό που μόλις σχημάτισε, το απελευθερώνει και στη συνέχεια δοκιμάζει ξανά.

Η DNA πολυμεράση λοιπόν είναι εκείνο το ένζυμο, το οποίο δημιουργεί τον νέο κλώνο στα εκμαγεία του διαχωρισμένου κλώνου από το DNA. Ωστόσο, όπως περιγράψαμε ανωτέρω η DNA πολυμεράση για να προσθέτει νουκλεοτίδια πρέπει να είναι συνδεδεμένη σε κάποιο ήδη ζευγαρωμένο νουκλεοτίδιο, να ελέγξει αν η συμπληρωματική βάση προστέθηκε σωστά και μετά να συνεχίσει στο επόμενο. Μετά τον διαχωρισμό όμως των δύο κλώνων του DNA δεν υπάρχει κανένα ζευγαρωμένο νουκλεοτίδιο ώστε να συνδεθεί η DNA πολυμεράση. Για το σκοπό αυτό απαιτείται ένα διαφορετικό ένζυμο η πριμάση, ικανό να αρχίζει μια νέα πολυνουκλεοτιδική αλυσίδα με απλή σύνδεση δύο νουκλεοτιδίων, χωρίς να χρειάζεται ένα ζευγαρωμένο άκρο. Το ένζυμο αυτό απλώς χρησιμοποιεί τον κλώνο εκμαγείο για να δημιουργήσει ένα μικρό κομμάτι ενός άλλου νουκλεϊκού οξέος που είναι γνωστό ως ριβονουκλεϊκό οξύ (ribonucleic acid - RNA). Το κομμάτι αυτό RNA, το οποίο περιέχει περίπου 10 νουκλεοτίδια, σχηματίζει ένα ζευγαρωμένο 3' άκρο ζευγαρώνοντας τα νουκλεοτίδια του με το εκμαγείο. Το σημείο αυτό πλέον μπορεί να αποτελέσει ένα σημείο έναρξης για την δράση της DNA πολυμεράση. Με βιολογικούς όρους λέμε ότι το RNA λειτουργεί ως εκκινητής (primer) για τη σύνθεση του DNA. Ένας κλώνος RNA από χημική άποψη μοιάζει πολύ με έναν μονό κλώνο DNA. Ωστόσο, οι επιμέρους μονάδες του είναι ριβονουκλεοτίδια, στα οποία το σάκχαρο είναι η ριβόζη και όχι η δεοξυριβόζη. Επίσης, διαφέρει από το DNA δεδομένου ότι αντί για θυμίνη (T) περιέχει ουρακίλη (U). Η ουρακίλη ζευγαρώνει με την αδενίνη και, έτσι, ο εκκινητής RNA συντίθεται πάνω στο εκμαγείο με βάση τους κανόνες συμπληρωματικότητας των βάσεων, ακριβώς όπως και ένας νέος κλώνος DNA. Το σχήμα 1.4 αντιπαραθέτει το μόριο του DNA με αυτό του RNA.

Δεδομένης της κατεύθυνσης, με την οποία η DNA πολυμεράση συνθέτει DNA στον προπορευόμενο κλώνο, ένας εκκινητής RNA χρειάζεται μόνο για την έναρξη της αντιγραφής σε μια αφετηρία αντιγραφής. Αντίθετα, στον καθυστερημένο κλώνο, όπου η σύνθεση του DNA είναι ασυνεχής, νέοι εκκινητές απαιτούνται συνεχώς. Καθώς μετακινείται η διχάλα αντιγραφής, αφήνει εκτεθειμένο ένα νέο τμήμα ασύζευκτων βάσεων. Έτσι κατά μήκος του καθυστερημένου κλώνου σχηματίζονται διάφοροι νέοι εκκινητές RNA. Η DNA πολυμεράση προσθέτει ένα δεοξυριβονουκλεοτίδιο στο 3' άκρο κάθε εκκινητή για να ξεκινήσει τη σύνθεση ενός κλώνου DNA, τον οποίο συνεχίζει να επιμηκύνει έως ότου συναντήσει τον επόμενο εκκινητή RNA.

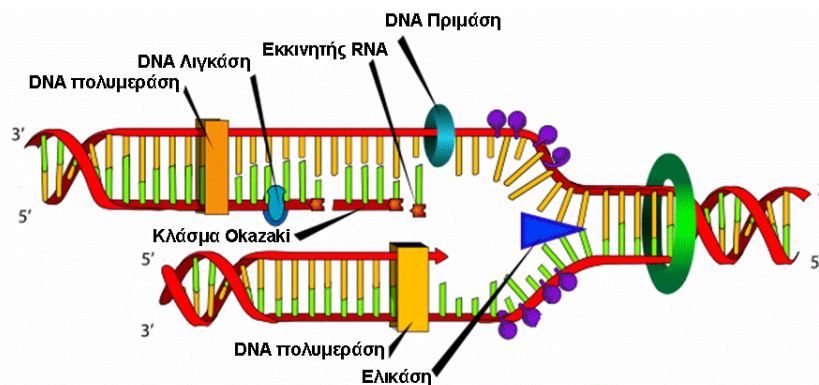


Σχήμα 1.4: Το Ριβονουκλεϊκό οξύ (RNA) σε σύγκριση με το δεοξυριβονουκλεϊκό οξύ (DNA).

Για την δημιουργία του καθυστερημένου κλώνου τρία ακόμα ένζυμα απαιτούνται με σκοπό να απομακρύνουν τους πολλούς εκκινητές και να τους αντικαταστήσουν με DNA, αλλά και για να ενώσουν τα επιμέρους τμήματα που δημιουργούνται από την ασυνεχή δημιουργία του και τα οποία είναι γνωστά ως κλάσματα του Οκαζάκι (Okazaki fragments) [15-16]. Συγκεκριμένα, μια νουκλεάση (nuclease) αποδομεί τον εκκινητή RNA, μια πολυμεράση επιδιόρθωσης (repair polymerase) αντικαθιστά το RNA από DNA και, τέλος, το ένζυμο λιγκάση του DNA (DNA ligase) συνδέει το 5'-φωσφορικό άκρο ενός νέου κλάσματος DNA με το 3'-υδροξυλικό άκρο του επόμενου κλάσματος.

Εκτός από τη DNA πολυμεράση, μία ομάδα από πολλές πρωτεΐνες εμπλέκονται στην αντιγραφή του DNA. Οι περισσότερες από τις πρωτεΐνες αυτές θεωρείται ότι συνδέονται σ' ένα μεγάλο πολυενζυμικό σύμπλοκο, το οποίο μετακινείται ως ενιαία μονάδα κατά μήκος του DNA, επιτρέποντας την σύνθεση και των δύο κλώνων με ένα συντονισμένο ρυθμό. Η κυριότερη πρωτεΐνη που πρέπει να αναφερθεί είναι η ελικάση (helicase), η οποία μετακινείται γρήγορα κατά μήκος του DNA, ξετυλίγοντας παράλληλα τη διπλή έλικα. Όπως είδαμε προηγουμένως, για να αρχίσει η αντιγραφή του DNA η διπλή έλικα πρέπει να είναι ανοικτή και να ανοίγει συνεχώς καθώς η διχάλα αντιγραφής προχωρεί, ώστε να

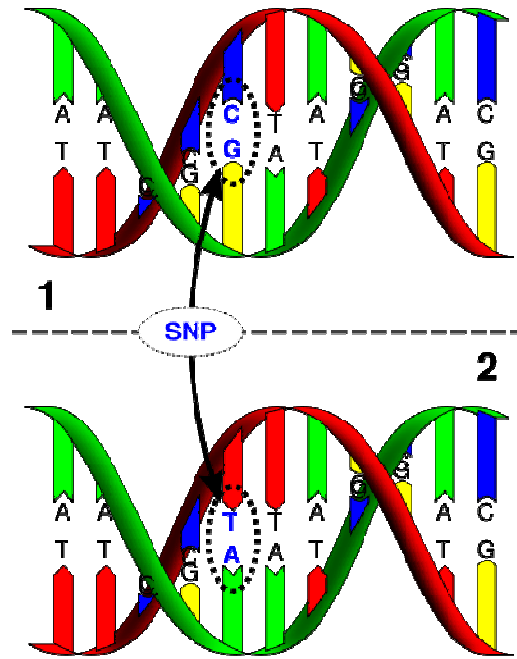
αποκαλύπτονται μονόκλωνα εκμαγεία για την πολυμεράση. Έχοντας παραθέσει όλη την απαιτούμενη γνώση για την διαδικασία της αντιγραφής του DNA, μπορούμε να την συνοψίσουμε στο σχήμα 1.5.



Σχήμα 1.5: Η διαδικασία της αντιγραφής.

1.4.Μεταλλάξεις και Απλοί Νουκλεοτιδικοί Πολυμορφισμοί

Η διαδικασία της αντιγραφής αποδεικνύεται να είναι πολύ ακριβής ώστε να μην επιτρέπει να συμβεί κάποια μόνιμη αλλαγή στο DNA του οργανισμού. Μια τέτοια μόνιμη αλλαγή αποκαλείται μετάλλαξη (mutation) και μπορεί να έχει σημαντικές επιπτώσεις στην ζωή του κυττάρου. Η τεχνολογία των μικροσυστοιχιών με την πάροδο του χρόνου επεκτάθηκε σε διάφορα πεδία της βιολογικής έρευνας. Ένα από αυτά είναι ο εντοπισμός των μεταλλάξεων στο DNA. Μια μετάλλαξη που επηρεάζει έστω και ένα ζεύγος νουκλεοτιδίων μπορεί να καταστρέψει έναν οργανισμό αν συμβεί σε μια ζωτική θέση στην αλληλουχία του DNA. Όταν η μετάλλαξη αφορά ένα μόνο νουκλεοτίδιο τότε ονομάζεται απλός νουκλεοτιδικός πολυμορφισμός (single-nucleotide polymorphism ή SNP). Έχει επικρατήσει με τον όρο SNP να εννοούμε τις μεταλλάξεις, οι οποίες αφορούν μόνο ένα νουκλεοτίδιο αλλά και που εμφανίζονται στο 1% και άνω του πληθυσμού. Οι απλοί νουκλεοτιδικοί πολυμορφισμοί λοιπόν έχουν οδηγήσει σε διάφορες παραλλαγές γονιδίων τα οποία ονομάζονται αλληλόμορφα (alleles). Για παράδειγμα εάν έχουμε δύο αλληλουχίες AAGCCTA και AAGCTTA οι οποίες βρίσκονται στο ίδιο κλάσμα ενός γονιδίου, τότε μπορούμε να πούμε ότι στην περίπτωση αυτή υπάρχει ένας απλός νουκλεοτιδικός πολυμορφισμός αλλά και δύο αλληλόμορφα του γονιδίου αυτού. Παρόμοιο παράδειγμα μπορούμε να δούμε σχηματικά στο σχήμα 1.6.



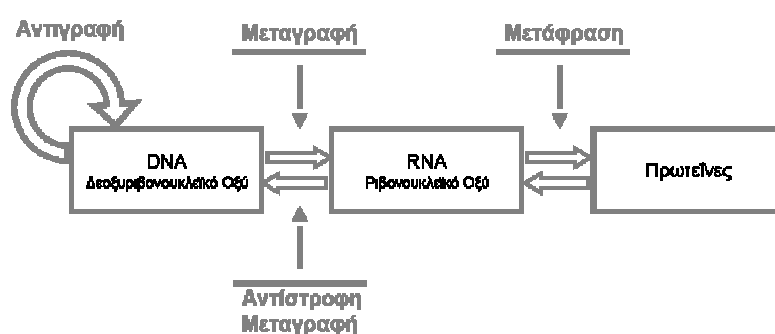
Σχήμα 1.6: Ένας απλός νουκλεοτιδικός πολυμορφισμός (SNP).

Η έρευνα τα τελευταία χρόνια έχει στραφεί προς τον εντοπισμό των απλών νουκλεοτιδικών πολυμορφισμών, αλλά και την ευθύνη που έχουν στην δημιουργία παθογενειών και ασθενειών. Όπως θα δούμε και στο επόμενο κεφάλαιο η εταιρία Illumina έχει ήδη κατασκευάσει μικροσυστοιχίες για αυτό τον σκοπό, όπως αυτές προέκυψαν στα πλαίσια ενός διεθνούς προγράμματος που ονομάζεται Harmap [17-18]. Η διαδικασία λοιπόν του εντοπισμού όλων αυτών των πολυμορφισμών που υπάρχουν στον πληθυσμό βρίσκεται ήδη σε καλό σημείο καθώς έχουν εντοπιστεί περί τα 5 από τα 10 εκατομμύρια, όπως υπολογίζονται. Αυτό το γεγονός όμως δεν σημαίνει ότι και όλοι αυτοί οι πολυμορφισμοί μπορεί να προκαλέσουν την εμφάνιση παθογενειών. Έτσι, πέραν του εντοπισμού, απαραίτητο είναι να επισημανθούν οι πολυμορφισμοί εκείνοι, οι οποίοι παίζουν κάποιο ρόλο στην εξέλιξη ασθενειών.

1.5.Γονιδιακή Έκφραση

Η έκφραση των γονιδίων [19-20] σε πρωτεΐνες είναι και το πεδίο με το οποίο ασχολήθηκαν τα πρώτα πειράματα μικροσυστοιχιών [21-25]. Όταν ένα κύτταρο του οργανισμού χρειαστεί κάποια πρωτεΐνη, το DNA δεν είναι αυτό που κατευθύνει την σύνθεση της. Η αλληλουχία των νουκλεοτιδίων του DNA που απαιτείται για την

δημιουργία της πρωτεΐνης αυτής αντιγράφεται σε μορφή RNA. Τα RNA αντίγραφα των τμημάτων του DNA χρησιμοποιούνται ως εκμαγεία που κατευθύνουν τη σύνθεση των πρωτεϊνών. Πιο γενικά δηλαδή, η ροή των γενετικών πληροφοριών στα κύτταρα γίνεται από το DNA στο RNA και από εκεί στις πρωτεΐνες. Αυτές είναι οι μορφές από τις οποίες διέρχονται οι γενετικές πληροφορίες κατά την έκφραση ενός γονιδίου, γεγονός το οποίο αποτελεί μια από τις πιο θεμελιώδεις αρχές της Βιολογίας. Δύο είναι οι βασικοί μηχανισμοί με τους οποίους οι γενετικές πληροφορίες μεταπηδούν από την μια μορφή στην άλλη, η μεταγραφή και η μετάφραση. Η έκφραση των γενετικών πληροφοριών συνοψίζονται στο σχήμα 1.7.



Σχήμα 1.7: Οι διαδικασίες με τις οποίες εκφράζεται η γενετική πληροφορία.

1.5.1.Μεταγραφή του DNA σε RNA

Η έκφραση ενός τμήματος των γενετικών οδηγιών ενός κυττάρου ξεκινά με την αντιγραφή της κατάλληλης αλληλουχίας των νουκλεοτιδίων του DNA, δηλαδή του κατάλληλου γονιδίου, σε μια αλληλουχία νουκλεοτιδίων του RNA. Αυτός ο μηχανισμός ονομάζεται μεταγραφή (transcription) επειδή παρόλο που οι πληροφορίες αντιγράφονται σε μια διαφορετική χημική μορφή, εξακολουθούν να γράφονται ουσιαστικά στην ίδια γλώσσα, δηλαδή στη γλώσσα των νουκλεοτιδίων. Όπως και το DNA, το RNA είναι ένα γραμμικό πολυμερές που αποτελείται από τέσσερα διαφορετικά νουκλεοτίδια, τα οποία συνδέονται μεταξύ τους με φωσφοδιεστερικούς δεσμούς. Όπως αναφέρθηκε και παραπάνω η δύο βασικές διαφορές του RNA από το DNA είναι το σάκχαρο, αλλά και η μία εκ των τεσσάρων βάσεων. Παρά το γεγονός ότι οι χημικές αυτές διαφορές δεν είναι μεγάλες, εν τούτοις το DNA και το RNA διαφέρουν πολύ ως προς τη δομή τους. Αφενός το DNA συναντάτε στα κύτταρα πάντα με την μορφή μιας δίκλωνης έλικας σε αντίθεση με το RNA το οποίο είναι μονόκλωνο. Αφετέρου, μια αλυσίδα RNA μπορεί να διπλωθεί σε ποικίλες

δομές, όπως διπλώνεται μια πολυπεπτιδική αλυσίδα για να σχηματίσει την τελική δομή μιας πρωτεΐνης, ή για να μεταφέρει πληροφορίες από το DNA στις πρωτεΐνες.

Η διαδικασία της μεταγραφής, είναι μια διεργασία παρόμοια με αυτήν της αντιγραφής του DNA. Και η μεταγραφή όπως και η αντιγραφή του DNA ξεκινάει με τον διαχωρισμό της διπλής έλικας σε κάποιο κομμάτι του DNA. Ακολουθώντας τους κανόνες της συμπληρωματικότητας των βάσεων, το RNA συντίθεται έχοντας ως εκμαγείο τον έναν κλώνο της διπλής έλικας. Όταν ένα εισερχόμενο ριβονουκλεοτίδιο ταιριάζει καλά με ένα δεοξυριβονουκλεοτίδιο του εκμαγείου, τότε θα συνδεθεί ομοιοπολικά με την αυξανόμενη αλυσίδα του RNA με μια ενζυμικά καταλυόμενη αντίδραση. Το παράγωγο RNA από την διαδικασία της μεταγραφής ονομάζεται μετάγραφο (transcript). Τα ένζυμα που χρησιμοποιούνται για τη μεταγραφή αποκαλούνται RNA πολυμεράσες (RNA polymerases). Αντίστοιχα με την DNA πολυμεράση που καταλύει την αντιγραφή του DNA, οι RNA πολυμεράσες καταλύουν το σχηματισμό των φωσφοδιεστερικών δεσμών που συνδέουν τα νουκλεοτίδια μεταξύ τους και δημιουργούν τον σακχαροφωσφορικό σκελετό της αλυσίδας του RNA. Με παρόμοιο τρόπο με την DNA πολυμεράση η RNA πολυμεράση μετακινείται βήμα-βήμα κατά μήκος του DNA, προχωρώντας μπροστά και ξετυλίγοντας την έλικα του DNA ώστε να αποκαλυφθεί μια νέα περιοχή του εκμαγείου κατάλληλη για ζευγάρωμα συμπληρωματικών βάσεων.

Εκτός από τις μεγάλες ομοιότητες της μεταγραφής με την αντιγραφή του DNA υπάρχουν και μεγάλες διαφορές μεταξύ των δύο αυτών μηχανισμών. Ο κλώνος του RNA που παράγεται από την διαδικασία της μεταγραφής δεν συνδέεται με το εκμαγείο όπως γίνεται στην αντιγραφή του DNA, αλλά πίσω από την περιοχή όπου προστέθηκαν τα ριβονουκλεοτίδια, η έλικα του DNA αναδιοργανώνεται και εκτοπίζει την αλυσίδα του RNA. Τα μόρια του RNA είναι πολύ μικρότερα του μορίου του DNA εφόσον αυτά παράγονται από ένα μικρό κομμάτι του DNA. Ενδεικτικά μπορεί κανείς να αναφέρει ότι το DNA ενός ανθρώπινου χρωμοσώματος μπορεί να περιέχει έως και 250 εκατομμύρια ζεύγη νουκλεοτιδίων, ενώ τα περισσότερα μόρια του RNA περιέχουν το μέγιστο λίγες χιλιάδες νουκλεοτίδια. Άλλες σημαντικές διαφορές έχουν να κάνουν με την κατάλυση των μηχανισμών από τις αντίστοιχες πολυμεράσες. Η RNA πολυμεράση καταλύει τη σύνδεση ριδονουκλεοτιδίων και όχι δεοξυριβονουκλεοτιδίων όπως η DNA πολυμεράση. Η RNA

πολυμεράσες δεν διαθέτουν τη δυνατότητα ελέγχου της πιστότητας του μεταγράφου, σε αντίθεση με την DNA πολυμεράση όπως είδαμε ανωτέρω. Αυτό έχει ως αποτέλεσμα το γεγονός ότι η RNA πολυμεράση δεν χρειάζεται εκκινήτες για να ξεκινήσει την μεταγραφή, ενώ προδίδει και το γεγονός ότι η μεταγραφή δεν χρειάζεται να είναι εξίσου ακριβής όσο η αντιγραφή του DNA, επειδή το RNA δεν χρησιμοποιείται για τη μόνιμη αποθήκευση των γενετικών πληροφοριών των κυττάρων.

Τα περισσότερα από τα γονίδια του DNA σκοπεύουν στον καθορισμό των αλληλουχιών των αμινοξέων των πρωτεϊνών που θα παραχθούν. Το RNA που μεταγράφεται από αυτά τα γονίδια ονομάζεται αγγελιοφόρο RNA (messenger RNA - mRNA). Στους ευκαριωτικούς οργανισμούς, ένα μόριο mRNA περιέχει πληροφορίες, οι οποίες έχουν μεταγραφεί από ένα μόνο γονίδιο και κωδικοποιεί μια μοναδική πρωτεΐνη. Αντιθέτως, στα βακτήρια, μια ομάδα από γειτονικά γονίδια μεταγράφονται σ' ένα ενιαίο mRNA, το οποίο όμως περιέχει τις πληροφορίες για αρκετές διαφορετικές πρωτεΐνες. Τα υπόλοιπα γονίδια παράγουν μη αγγελιοφόρα μόρια RNA, τα οποία χρησιμεύουν ως δομικά και ενζυμικά συστατικά των κυττάρων και παίζουν καίριο ρόλο στη μετάφραση του γενετικού μηνύματος σε πρωτεΐνη. Δύο άλλα είδη RNA είναι το ριβοσωματικό RNA (ribosomal RNA - rRNA), το οποίο σχηματίζει το κέντρο των ριβοσωμάτων, πάνω στα οποία πραγματοποιείται η μετάφραση του mRNA σε πρωτεΐνη, και το μεταφορικό RNA (transfer RNA - tRNA), το οποίο σχηματίζει τους συναρμολογητές που επιλέγουν τα αμινοξέα και τα συγκρατούν στην κατάλληλη θέση πάνω στο ριβοσωμάτιο για να ενσωματωθούν σε μια πρωτεΐνη.

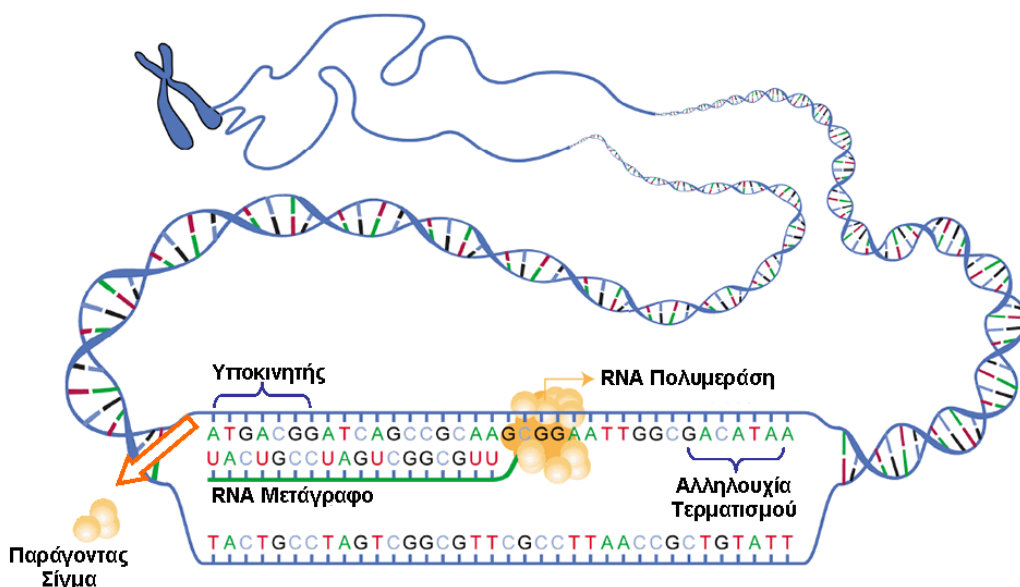
Δεδομένου ότι μόνο μικρά κομμάτια του DNA μεταγράφονται σε RNA, με κάποιο τρόπο θα πρέπει να υποδεικνύεται στην RNA πολυμεράση από ποιο σημείο του DNA θα ξεκινάει την μεταγραφή και σε ποιο σημείο θα σταματάει. Ένα μόριο RNA πολυμεράσης τείνει να προσκολληθεί πάνω στο DNA και στη συνέχεια να μετακινηθεί κατά μήκος του DNA μέχρι να εντοπίσει μια περιοχή, η οποία ονομάζεται υποκινητής (promoter). Ο παράγοντας σίγμα (σ) (sigma factor), που περιέχεται στην RNA πολυμεράση είναι υπεύθυνος για τον εντοπισμό της αλληλουχίας του υποκινητή πάνω στο DNA. Όταν η πολυμεράση προσδεθεί στον υποκινητή και συνθέσει ένα τμήμα RNA μήκους περίπου 10 νουκλεοτιδίων, ο παράγοντας σίγμα απελευθερώνεται, οπότε το ένζυμο είναι πλέον

ελεύθερο να μετακινηθεί προς τα εμπρός και να συνεχίσει τη μεταγραφή χωρίς τον παράγοντα. Μετά την απελευθέρωση από την αλληλουχία τερματισμού του DNA, η RNA πολυμεράση επανασυνδέεται με έναν ελεύθερο παράγοντα σίγμα και αναζητά έναν υποκινητή, όπου μπορεί να αρχίσει ξανά τη διεργασία της μεταγραφής. Όταν η RNA πολυμεράση εντοπίσει τον υποκινητή, ο οποίος περιέχει μια αλληλουχία νουκλεοτιδίων που σηματοδοτεί τη θέση έναρξης για τη σύνθεση του RNA, προσδένεται ισχυρά επάνω στο μόριο του DNA. Στην συνέχεια η RNA πολυμεράση διανοίγει την περιοχή της διπλής έλικας που βρίσκεται ακριβώς μπροστά της και έτσι αποκαλύπτει τα νουκλεοτίδια σ' ένα βραχύ τμήμα του DNA κάθε κλώνου. Ένας από τους κλώνους DNA δρα ως εκμαγείο ώστε να τοποθετηθούν σωστά οι συμπληρωματικές βάσεις. Η διαδικασία συνεχίζεται έως ότου η RNA πολυμεράση συναντήσει ένα δεύτερο σήμα πάνω στο DNA, την αλληλουχία τερματισμού (stop site), οπότε σταματά και απελευθερώνει τόσο το DNA εκμαγείο όσο και τη νεοσυνθετημένη αλυσίδα του RNA. Τα παραπάνω δίνουν μια γενική περιγραφή του μηχανισμού της μεταγραφής στους προκαρυωτικούς οργανισμούς. Ο μηχανισμός παρουσιάζεται και στο σχήμα 1.8.

Υπάρχει διαφοροποίηση του μηχανισμού της μεταγραφής από τους προκαρυωτικούς στους ευκαρυωτικούς οργανισμούς. Το βακτηριακό DNA είναι άμεσα εκτεθειμένο στο κυτταρόπλασμα, το οποίο περιέχει τα ριβοσωμάτια πάνω στα οποία επιτελείται η σύνθεση των πρωτεϊνών. Αντίθετα, στα ευκαρυωτικά κύτταρα, το DNA περικλείεται στον πυρήνα. Η μεταγραφή γίνεται στον πυρήνα ενώ η πρωτεϊνοσύνθεση στα ριβοσωμάτια δηλαδή στο κυτταρόπλασμα. Επομένως, για να μεταφρασθεί ένα ευκαρυωτικό mRNA πρέπει προηγουμένως να μεταφερθεί έξω από τον πυρήνα διαμέσου των μικρών πόρων του πυρηνικού περιβλήματος.

Έτσι το πρωτογενές μετάγραφο (primary transcript) που παράγεται από την μεταγραφή στον πυρήνα, υφίσταται κάποια επεξεργασία ώστε στην συνέχεια να διοχετευθεί στο κυτταρόπλασμα όπου θα γίνει η μετάφραση. Η επεξεργασία αυτή για τα πρωτογενή μετάγραφα που προορίζονται να γίνουν μόρια mRNA, περιλαμβάνει τον σχηματισμό της καλύπτρας του RNA (RNA capping) και την πολυαδενυλίωση (polyadenylation). Οι δύο αυτές τροποποιήσεις αφενός αυξάνουν τη σταθερότητα των μορίων του mRNA αφετέρου χρησιμοποιούνται από τον πρωτεϊνοσυνθετικό μηχανισμό ως ένδειξη ότι και τα δύο άκρα

του mRNA είναι παρόντα και, επομένως, ότι το μήνυμα είναι πλήρες. Η κάθε μία από αυτές της τροποποιήσεις ουσιαστικά οριοθετεί το ένα από τα δύο άκρα της αλληλουχίας.



Σχήμα 1.8: Ο μηχανισμός της μεταγραφής.

1.5.2. Αντίστροφη μεταγραφή

Η αντίστροφη μεταγραφή είναι μια διεργασία, η οποία χρησιμοποιείται ευρέως στην τεχνολογία των μικροσυστοιχιών. Μέσω αυτής της διεργασίας παράγεται το συμπληρωματικό DNA με χρήση του οποίου δημιουργούνται οι αντίστοιχες cDNA μικροσυστοιχίες. Σύμφωνα με το κεντρικό δόγμα της βιολογίας, οι πρωτεΐνες παράγονται έπειτα από μεταγραφή του DNA σε mRNA. Τα γονίδια στα ευκαρυωτικά κύτταρα συνήθως περιέχουν εσόνια σε αντίθεση με τα γονίδια στα προκαρυωτικά κύτταρα. Τα εσόνια (ή αλλιώς ιντρόνια) δεν μεταφράζονται σε αμινοξέα, και για τον λόγο αυτόν πρέπει να αφαιρεθούν από το mRNA πριν από την μετάφραση.

Ορισμένες φορές για παράδειγμα, χρειάζεται να εκφραστούν ευκαρυωτικά γονίδια από προκαρυωτικά κύτταρα. Αυτό συνήθως γίνεται με την μεταφορά του DNA ενός ευκαρυωτικού κυττάρου σε ένα προκαρυωτικό κύτταρο-ξενιστή, το οποίο θα μεταγραφεί σε mRNA και θα μεταφραστεί σε πρωτεΐνη. Τα προκαρυωτικά κύτταρα, όπως αναφέρθηκε όμως δεν έχουν τη δυνατότητα να αφαιρέσουν τα εσόνια από το πρόδρομο mRNA,

επομένως η αφαίρεσή τους πρέπει να γίνει πριν την μεταφορά. Για τον λόγο αυτόν από το mRNA παράγεται το συμπληρωματικό DNA, το οποίο δεν περιέχει εσώνια.

Το cDNA είναι ένα μόριο, το οποίο έχει υποστεί τις απαραίτητες αποκοπές και επεξεργασίες, ώστε να είναι έτοιμο προς μετάφραση. Για την σύνθεση του cDNA χρησιμοποιείται το ένζυμο, την αντίστροφη μεταγραφάση. Το cDNA χρησιμοποιείται ευρύτατα για την κλωνοποίηση γονιδίων ευκαρυωτικών κυττάρων σε προκαρυωτικά, ώστε να παραχθεί η πρωτεΐνη, που αυτά κωδικοποιούν, σε μεγάλες ποσότητες. Κάθε προκαρυωτικό κύτταρο που δέχεται το γονίδιο πολλαπλασιάζεται, παράγοντας έτσι ένα βακτηριακό κλώνο. Το σύνολο των βακτηριακών κλώνων που περιέχουν αντίγραφα όλων των ώριμων mRNA των γονιδίων που εκφράζονται σε έναν κυτταρικό τύπο ονομάζεται cDNA βιβλιοθήκη.

1.5.3.Μετάφραση

Δεδομένου ότι υπάρχουν και πρωτεϊνικές μικροσυστοιχίες καλό είναι να κάνουμε μια μικρή αναφορά στον τρόπο με τον οποίο τελικά μεταφράζεται η γενετική πληροφορία από το RNA σε πρωτεΐνη. Οι ερευνητές είχαν ήδη εξάγει το γεγονός ότι οι πληροφορίες που κωδικοποιούνται στο DNA αντιγράφονται αρχικά σε RNA και κατόπιν σε πρωτεΐνες, πριν από τα μέσα του 20^{ου} αιώνα. Ωστόσο, εκκρεμούσε ακόμα το ερώτημα, πώς οι πληροφορίες που περιέχονται στη γραμμική αλληλουχία του RNA μεταφράζονται σε μια γραμμική αλληλουχία ενός πολύ διαφορετικού από χημικής άποψης συνόλου μονάδων, δηλαδή τα αμινοξέα στις πρωτεΐνες.

Το παραγόμενο από τον μηχανισμό της μεταγραφής mRNA περιέχει όλες τις απαραίτητες πληροφορίες στην αλληλουχία των νουκλεοτιδίων του για την σύνθεση μιας πρωτεΐνης. Η μετατροπή των πληροφοριών του RNA σε πρωτεΐνη αντιπροσωπεύει μια μετάφραση των πληροφοριών σε μια άλλη γλώσσα που χρησιμοποιεί πολύ διαφορετικά σύμβολα. Η δυσκολία στην κατανόηση της μετάφρασης έγκειται στο γεγονός ότι στο mRNA οι δομικές μονάδες είναι τα τέσσερα διαφορετικά νουκλεοτίδια, ενώ στις πρωτεΐνες υπάρχουν είκοσι διαφορετικά αμινοξέα. Στις αρχές της δεκαετίας του 1960 όμως

αποκρυπτογραφήθηκε η αντιστοιχία των αμινοξέων των πρωτεϊνών με τα νουκλεοτίδια του mRNA. Αντιστοιχία αυτή ονομάζεται γενετικός κώδικας (genetic code) [26-27] και παρουσιάζεται στον πίνακα 1.1.

Πίνακας 1.1: Ο γενετικός κώδικας.

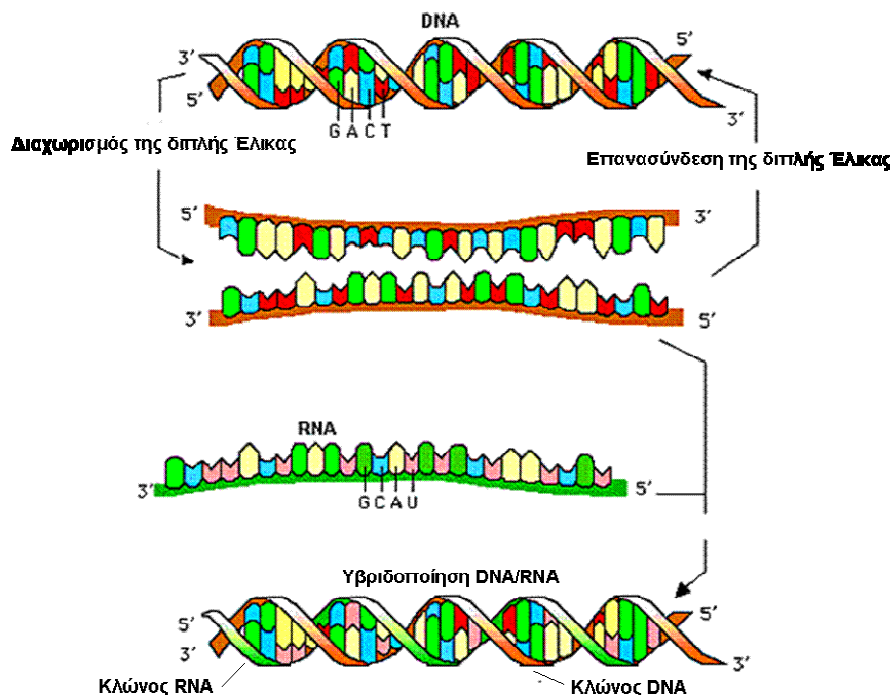
| | U | C | A | G | |
|---|-----|-----|------|------|---|
| U | Phe | Ser | Tyr | Cys | U |
| | Phe | Ser | Tyr | Cys | C |
| | Leu | Ser | STOP | STOP | A |
| | Leu | Ser | STOP | Trp | G |
| C | Leu | Pro | His | Arg | U |
| | Leu | Pro | His | Arg | C |
| | Leu | Pro | Gln | Arg | A |
| | Leu | Pro | Gln | Arg | G |
| A | Ile | Thr | Asn | Ser | U |
| | Ile | Thr | Asn | Ser | C |
| | Ile | Thr | Lys | Arg | A |
| | Met | Thr | Lys | Arg | G |
| G | Val | Ala | Asp | Gly | U |
| | Val | Ala | Asp | Gly | C |
| | Val | Ala | Glu | Gly | A |
| | Val | Ala | Glu | Gly | G |

Για την δημιουργία των πρωτεϊνών η αλληλουχία των νουκλεοτιδίων ενός μορίου mRNA διαβάζεται διαδοχικά ανά ομάδες τριών νουκλεοτιδίων. Οι δυνατοί συνδυασμοί τριών νουκλεοτιδίων είναι $4 \times 4 \times 4 = 64$. Ωστόσο, στις πρωτεΐνες υπάρχουν συνήθως μόνο 20 διαφορετικά αμινοξέα. Επομένως, είτε ορισμένες τριπλέτες νουκλεοτιδίων δεν χρησιμοποιούνται ποτέ είτε ορισμένα αμινοξέα καθορίζονται από περισσότερες από μια τριπλέτες, όπως φαίνεται και στον πίνακα 1.1. Κάθε ομάδα τριών διαδοχικών νουκλεοτιδίων του RNA αποκαλείται κωδικόνιο (codon) και κάθε κωδικόνιο καθορίζει ένα αμινοξύ.

1.6. Υβριδοποίηση νουκλεϊκών οξέων

Η υβριδοποίηση [28-29] των νουκλεϊκών αλληλουχιών είναι η πιο σημαντική ιδιότητα, την οποία εκμεταλλεύεται η τεχνολογία των μικροσυστοιχιών. Εάν ένα γονίδιο απομονωθεί από το γονιδίωμα σαν ένα κομμάτι DNA είναι σημαντικό να μπορούσαμε να

γνωρίζουμε το χρωμόσωμα από το οποίο προήλθε, καθώς επίσης και την θέση του πάνω στο χρωμόσωμα. Επίσης είναι σημαντικό να μπορούσαμε να γνωρίζουμε σε ποια κύτταρα του οργανισμού μεταγράφεται το συγκεκριμένο γονίδιο, ή αν υπάρχουν κύτταρα και άλλων οργανισμών που διαθέτουν παρόμοια γονίδια. Εργαστηριακά όλα τα παραπάνω υλοποιήθηκαν με την χρήση της ιδιότητας της υβριδοποίησης, πάνω στην οποία στηρίχθηκαν και τα παράλληλα πολλαπλά πειράματα που διενεργούνται σε μια μικροσυστοιχία. Σύμφωνα με την ιδιότητα αυτή ένας κλώνος DNA μπορεί να ζευγαρώσει εκλεκτικά με έναν δεύτερο κλώνο, με συμπληρωματική αλληλουχία νουκλεοτιδίων. Σε προηγούμενα εδάφια συζητήθηκε το γεγονός ότι οι δύο κλώνοι μιας διπλής έλικας DNA συνδέονται με σχετικά ασθενείς δεσμούς υδρογόνου, οι οποίοι μπορούν να διασπαστούν είτε με θέρμανση του DNA είτε με έκθεση του σε διαλύματα με ακραίες τιμές pH. Με τον τρόπο αυτόν διαχωρίζονται οι δύο κλώνοι του DNA χωρίς όμως να διασπών τους ομοιοπολικούς δεσμούς που συνδέουν τα νουκλεοτίδια μεταξύ τους σε κάθε κλώνο. Η λογική της υβριδοποίησης ακολουθεί ακριβώς την αντίστροφη διαδικασία. Αν δηλαδή η θερμοκρασία ελαττωθεί σταδιακά στα φυσιολογικά επίπεδα ή αν το pH επανέλθει σε ουδέτερη τιμή, οι συμπληρωματικοί κλώνοι θα δημιουργήσουν πάλι διπλές έλικες, αποκαθιστώντας τους συμπληρωματικούς δεσμούς υδρογόνου. Η υβριδοποίηση παρουσιάζεται στο σχήμα 1.9.



Σχήμα 1.9: Η διεργασία της υβριδοποίησης του DNA.

Η υβριδοποίηση συμβαίνει ανάμεσα σε δύο οποιεσδήποτε μονόκλωνες αλυσίδες νουκλεοτιδικών οξέων (DNA/DNA, RNA/RNA ή RNA/DNA) με την προϋπόθεση ότι θα έχουν συμπληρωματική αλληλουχία νουκλεοτιδίων. Η ικανότητα ενός μονόκλωνου μορίου νουκλεϊκού οξέος να σχηματίζει διπλή έλικα μόνο με ένα συμπληρωματικό είναι και αυτή, η οποία αξιοποιείται από την τεχνολογία των μικροσυστοιχιών.

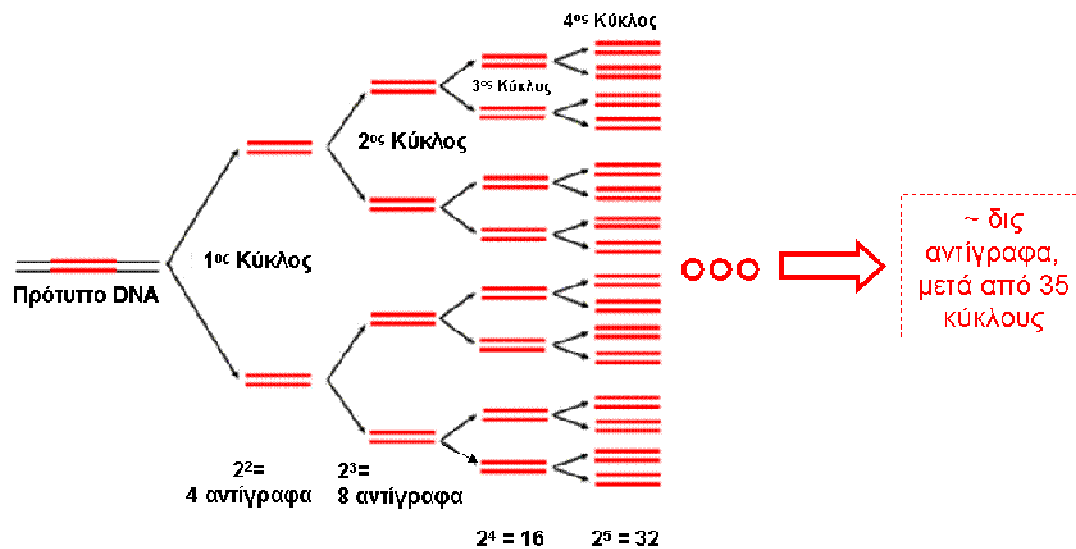
Η υβριδοποίηση χρησιμοποιείται ουσιαστικά για τον εντοπισμό συγκεκριμένων αλληλουχιών σε δείγματα. Για τον σκοπό αυτό είναι απαραίτητο να απομονωθεί μια αλληλουχία γνωστή ως ανιχνευτής (probe). Ένας DNA ανιχνευτής είναι ένα βραχύ, μονόκλωνο μόριο, με μήκος που κυμαίνεται συνήθως από 10 έως 1000 νουκλεοτίδια, το οποίο χρησιμοποιείται σε αντιδράσεις υβριδισμού για την ανίχνευση μορίων νουκλεϊκών οξέων με συμπληρωματική αλληλουχία. Οι ανιχνευτές πλέον παράγονται στο εργαστήριο με μη ενζυματικές μεθόδους, σε αντίθεση με τα προηγούμενα χρόνια όπου λαμβάνονταν μόνο από φυσικές πηγές.

1.7.Αλυσιδωτή Αντίδραση Πολυμεράσης

Η μεγάλες ποσότητες των δειγμάτων που χρειάζονται για ένα πείραμα μικροσυστοιχιών απαιτούν μια γρήγορη και αξιόπιστη τεχνική. Η διαδικασία που ακολουθούνταν παλαιότερα για την απομόνωση ενός γονιδίου ήταν η κλωνοποίηση μέσω DNA βιβλιοθηκών. Αυτό συνέβαινε έως ότου αναπτύχθηκε η τεχνική της Αλυσιδωτής Αντίδραση Πολυμεράσης (Polymerase Chain Reaction – PCR) [30-36], η οποία είναι μια ταχύτερη, φθηνότερη και πιο αποδοτική τεχνική που είχε πολλές εφαρμογές. Με την χρήση της PCR μια αλληλουχία νουκλεοτιδίων αντιγράφεται γρήγορα και αξιόπιστα σε μεγάλες ποσότητες, *in vitro* και χωρίς να χρειάζονται κύτταρα.

Η όλη διαδικασία στηρίζεται στην χρήση της DNA πολυμεράσης, η οποία δρα επαναληπτικά για την δημιουργία πολλών αντιγράφων. Μικρά κομμάτια της αλληλουχίας των νουκλεοτιδίων λειτουργούν ως εκκινητές (primers), τα οποία υβριδοποιούνται με το εκμαγείο, καθοδηγώντας έτσι την DNA πολυμεράση. Ο σχεδιασμός των εκκινητών γίνεται με τέτοιο τρόπο ώστε να είναι δυνατή η εκκίνηση της αντιγραφής σε κάθε κλώνο της

διπλής έλικας. Δεδομένου ότι οι εκκινητές παρασκευάζονται με χημική σύνθεση, η τεχνική PCR χρησιμοποιείται μόνο για την κλωνοποίηση κλασμάτων DNA με γνωστή αλληλουχία των δύο άκρων τους. Μετά από μερικές δεκάδες κύκλους της PCR, η DNA πολυμεράση παράγει έναν πολύ μεγάλο αριθμό αντιγράφων όπως ακριβώς φαίνεται στο σχήμα 1.10.



Σχήμα 1.10: Η αλυσιδωτή Αντίδραση πολυμεράσης.

Αρχικά σχεδιάζονται και συντίθεται τα δύο ολιγονουκλεοτίδια που θα δράσουν ως εκκινητές, ανάλογα με την αλληλουχία η οποία θα αντιγραφεί. Ο κάθε ένας από τους εκκινητές αυτούς είναι συμπληρωματικός με την αλληλουχία ενός κλώνου της δίκλωνης έλικας του DNA στα δύο αντίθετα άκρα της επιλεγμένης περιοχής. Με την βοήθεια των εκκινητών θα ξεκινήσει η *in vitro* σύνθεση του DNA από την DNA πολυμεράση. Όπως φαίνεται και στο σχήμα 1.10 η αλυσιδωτή αντίδραση πολυμεράσης πραγματοποιείται επαναληπτικά σε πολλούς κύκλους. Ο πρώτος κύκλος ξεκινάει με ένα δίκλωνο μόριο DNA, το οποίο θερμαίνεται με σκοπό να διαχωριστούν οι δύο κλώνοι. Μετά τον διαχωρισμό των κλώνων η θερμοκρασία του μείγματος της αντίδρασης ελαττώνεται, οπότε οι εκκινητές που βρίσκονται σε μεγάλη περίσσεια μπορεί να υβριδοποιηθούν με τις συμπληρωματικές προς αυτούς αλληλουχίες στους δύο κλώνους του DNA. Οι επόμενοι κύκλοι ξεκινούν ξανά με θέρμανση έτσι ώστε να διαχωριστούν οι νεοσυντεθειμένοι κλώνοι του DNA και τα βήματα ακολουθούνται όπως ακριβώς και κατά τον πρώτο κύκλο. Στους επόμενους κύκλους επίσης οι παραγόμενοι κλώνοι λειτουργούν και αυτοί με τη σειρά τους ως εκμαγεία. Είναι σημαντικό να αναφερθεί ότι γίνεται χρήση μιας ειδικής

DNA πολυμεράσης, η οποία είναι εξαιρετικά σταθερή και ανθεκτική στην αύξηση της θερμοκρασίας κατά την οποία συντελείτε ο διαχωρισμός της δίκλωνης αλυσίδας του DNA. Με την χρήση της ειδικής αυτής πολυμεράσης δεν απαιτείται η συνεχής πρόσθεση νέας πολυμεράσης στο μείγμα σε κάθε νέο κύκλο, επομένως, δεν είναι απαραίτητο να προστίθεται ξανά μετά από κάθε κύκλο της αντίδρασης. Τα επιθυμητά ποσά του DNA που απαιτούνται επιτυγχάνονται σε περίπου 30 κύκλος. Ο κάθε κύκλος διπλασιάζει την ποσότητα του DNA που είχε συντεθεί στον προηγούμενο κύκλο. Όσον αφορά την ταχύτητα της PCR, με την αυτοματοποίηση της επαναληπτικής διαδικασίας επιτυγχάνεται η κλωνοποίηση ενός κλάσματος του DNA μέσα σε λίγες ώρες χωρίς την παρουσία κυττάρων, σε αντίθεση με τις παραδοσιακές τεχνικές κλωνοποίησης που απαιτούσαν αρκετές ημέρες.

1.8.Συμπεράσματα

Στο κεφάλαιο αυτό ανατρέξαμε στις βασικές αρχές της βιολογίας. Συγκεκριμένα αναπτύχθηκαν οι αρχές που διέπουν το DNA, η διεργασίες της αντιγραφής του και της έκφρασης του σε πρωτεΐνες. Γίνεται εύκολα αντιληπτό ότι ένα πείραμα μικροσυστοιχιών είναι βασισμένο στις θεμελιώδης αρχές της βιολογίας, και όχι σε περίπλοκες διαδικασίες. Η παράλληλη όμως πειραματική διαδικασία, η οποία μπορεί να εξετάσει ένα μεγάλο πλήθος γονιδίων και άλλων βιολογικών αλληλουχιών, είναι και αυτή που προσδίδει τεράστια σημασία στην τεχνολογία αυτή. Το ανθρώπινο γονιδίωμα αποτελείται από εκατομμύρια γονίδια και για τον λόγο αυτό η μεμονωμένη διερεύνηση γονιδίων δεν μπορεί να εξάγει χρήσιμα συμπεράσματα. Δεδομένου ότι πλέον η τεχνολογία των μικροσυστοιχιών έχει καταφέρει την κατασκευή πλακιδίων ολόκληρου του ανθρώπινου γονιδιώματος, αναμένεται τα προσεχή χρόνια η συνεισφορά των μικροσυστοιχιών στην βιολογική έρευνα να αυξηθεί σημαντικά.

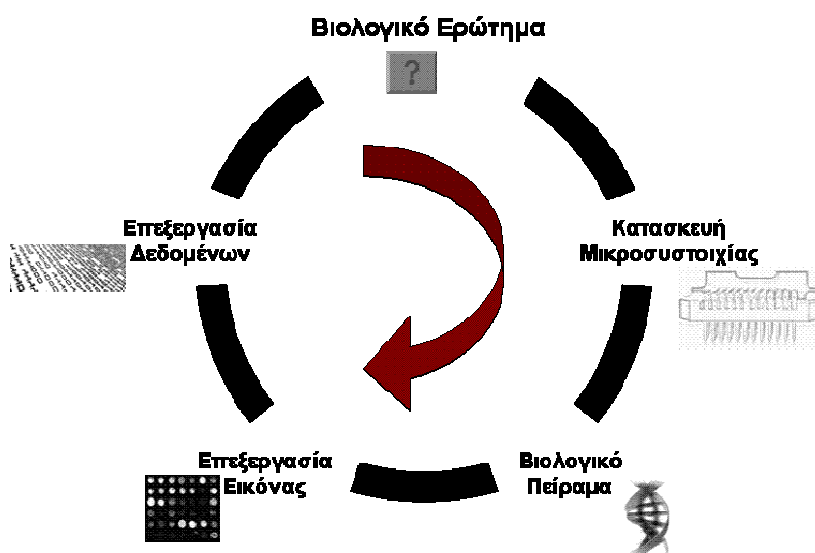
ΚΕΦΑΛΑΙΟ 2^ο: Εισαγωγή στις Μικροσυστοιχίες

2.Εισαγωγή στις μικροσυστοιχίες

2.1. Ιστορική αναδρομή

Η τεχνολογία μικροσυστοιχιών είναι μια τεχνολογία, η οποία εξελίσσεται με γοργούς ρυθμούς τα τελευταία δέκα με δεκαπέντε χρόνια και παρέχει το μέσο ώστε να επιτύχουμε με ταχύτητα παράλληλα πειράματα υβριδοποίησης. Με τη υπάρχουσα τεχνολογία, ένα πείραμα υβριδοποίησης μπορεί να παράγει τα πρότυπα έκφρασης εκατοντάδων χιλιάδων γονιδίων ταυτόχρονα. Ένα πείραμα μικροσυστοιχιών, όπως και τα περισσότερα πειράματα γενετικής, περιλαμβάνει συνήθως την απόκτηση και την επικύρωση μεγάλου συνόλου δεδομένων. Τα σύνολα δεδομένων αυτά περιέχουν μία ποικιλία από διαφορετικές πληροφορίες, από τις αλληλουχίες των γονιδίων ή των κλώνων που τοποθετούνται σε μια μικροσυστοιχία, έως τις ποσοτικοποιημένες τιμές έκφρασης για κάθε γονίδιο κάτω από διαφορετικές πειραματικές συνθήκες. Η διαδικασία ενός πειράματος μικροσυστοιχιών αρχίζει από το βιολογικό του σχεδιασμό και την κατασκευή της συστοιχίας, συνεχίζεται με την λήψη της εικόνας και την επεξεργασία της, και τελειώνει με την συγκέντρωση των

δεδομένων και την ανάλυση τους. Η ανάλυση των δεδομένων θα οδηγήσει πιθανώς σε μια νέα βιολογική υπόθεση, η οποία θα απαιτήσει στη συνέχεια επανάληψη του πειράματος, νέα κατασκευή της συστοιχίας κ.ο.κ. Ο παραπάνω κύκλος διεργασιών έχει ονομαστεί κύκλος ζωής των μικροσυστοιχιών, και φαίνεται στο σχήμα 1.



Σχήμα 2.1: Ο κύκλος ζωής των μικροσυστοιχιών

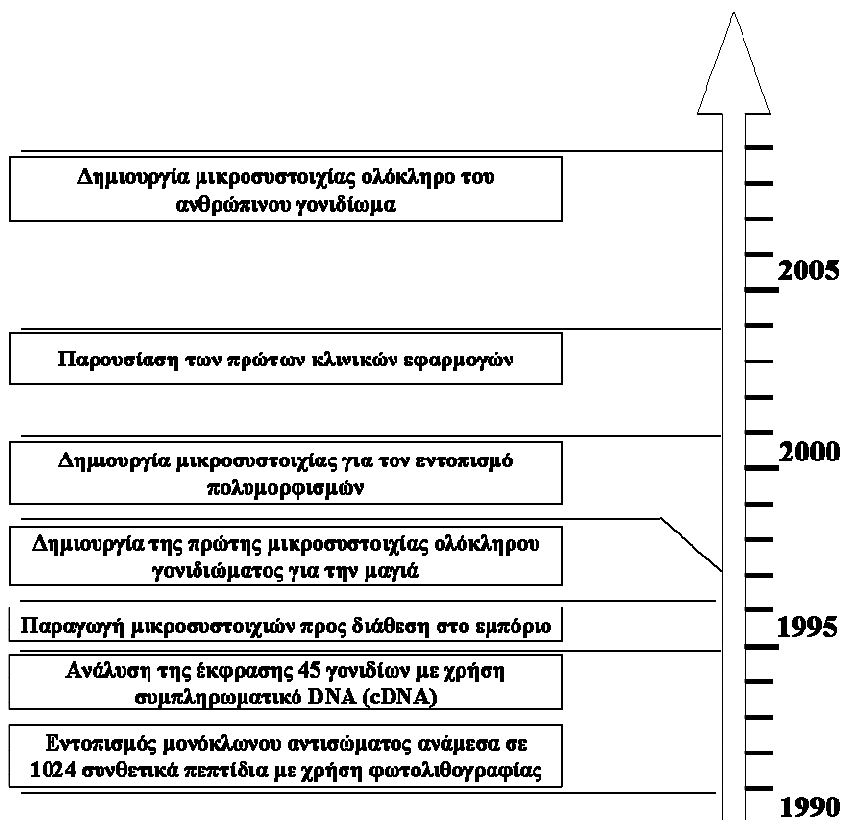
Κατά τη διάρκεια κάθε επανάληψης του κύκλου ζωής των μικροσυστοιχιών, παράγονται νέα στοιχεία, τα οποία μπορούν στη συνέχεια να χρησιμοποιηθούν είτε σε επόμενες επαναλήψεις, είτε στα επόμενα βήματα της ίδια επανάληψης. Παραδείγματος χάριν, τα δεδομένα που συνδέονται με τη διαδικασία κατασκευής της μικροσυστοιχίας, τα οποία περιέχουν πληροφορίες για τις θέσεις των αλληλουχιών πάνω στην μικροσυστοιχία, μπορούν να χρησιμοποιηθούν στην επεξεργασία εικόνας για να αυτοματοποιήσουν πολλές πτυχές της λειτουργίας. Ωστόσο, οι μέθοδοι οι οποίες προτείνονται για την επεξεργασία εικόνας στην παρούσα εργασία είναι αυτόματες και δεν απαιτούν καμιά εκ των προτέρων γνώση που να πηγάζει από προηγούμενα βήματα. Λόγω επίσης του μεγάλου όγκου των δεδομένων που παράγονται στα πειράματα μικροσυστοιχιών, η αρωγή ευφών πληροφοριακών συστημάτων καθίσταται απαραίτητη.

Δυο είναι οι βασικές απαιτήσεις που μπορούν να συμβάλουν στην περαιτέρω εξέλιξη των μικροσυστοιχιών. Αφενός πρέπει να υπάρχει συνεχής εξέλιξη του υλικού (Hardware) για την κατασκευή της μικροσυστοιχίας, αφετέρου να υλοποιηθούν τα υπολογιστικά εργαλεία για την ανάλυση του πλήθους των δεδομένων που θα παραχθούν, ώστε αυτά να

συμβάλουν στην κατανόηση των λειτουργιών των βιολογικών συστημάτων. Με τη σταθερή πρόοδο στην ανάπτυξη του υλικού των μικροσυστοιχιών, ο διαθέσιμος σήμερα εξοπλισμός μπορεί να παράγει μικροσυστοιχίες με εκατοντάδες χιλιάδες βιολογικές αλληλουχίες. Όσον αφορά στην ανάπτυξη των κατάλληλων εργαλείων για την αξιοποίηση του μεγάλου όγκου δεδομένων, υπάρχουν δύο σημαντικά ζητήματα, με τα οποία η επιστήμη της πληροφορικής συνεισφέρει. Το πρώτο έχει να κάνει με την εξαγωγή των ποσοτικοποιημένων τιμών από την εικόνα. Η διαδικασία αυτή όπως θα δούμε εκτενώς στην παρούσα διατριβή λαμβάνει χώρα με την χρήση τεχνικών επεξεργασίας εικόνας. Το δεύτερο και εξίσου σημαντικό ζήτημα είναι η επεξεργασία των εξαγόμενων από την εικόνα δεδομένων, με σκοπό την εξόρυξη νέας γνώσης. Η παρούσα διδακτορική διατριβή ασχολείται με το σχεδιασμό και την υλοποίηση ευφυών πληροφοριακών συστημάτων για την επεξεργασία των εικόνων των μικροσυστοιχιών.

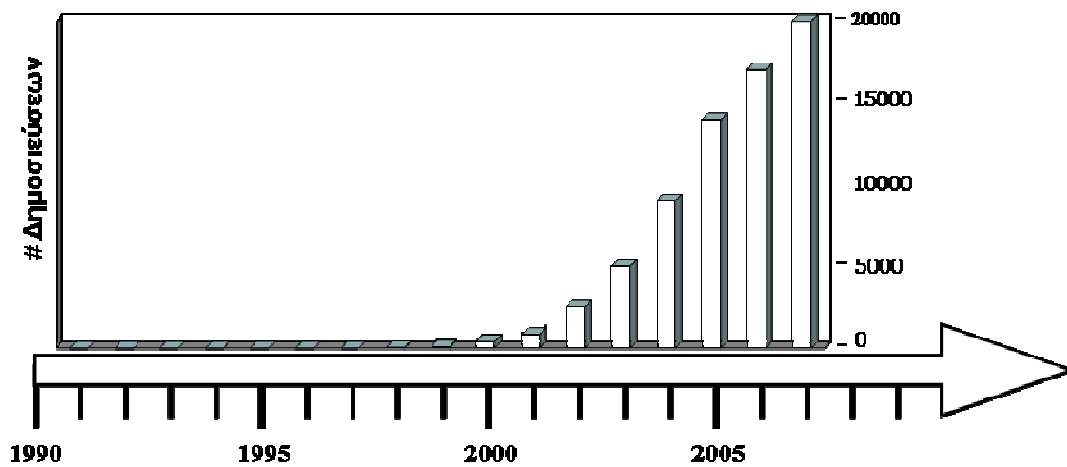
Ως προπομπός της τεχνολογίας αυτής μπορεί να χαρακτηριστεί η εργασία των Fodor και των συνεργατών του [37] το 1991, οι οποίοι χρησιμοποίησαν την φωτολιθογραφία με σκοπό να εντοπίσουν ένα μονόκλωνο αντίσωμα ανάμεσα σε 1024 συνθετικά πεπτίδια. Η πρώτη ωστόσο εργασία μικροσυστοιχιών παρουσιάστηκε το 1995 από τους Schena *et al.* [38], οι οποίοι ανέλυσαν την έκφραση 45 γονιδίων κάνοντας χρήση συμπληρωματικό DNA (cDNA). Η ανάλυση της έκφρασης των γονιδίων ήταν και η πρώτη εφαρμογή της τεχνολογίας αυτής. Ένα καθοριστικό βήμα για την διάδοση της τεχνολογίας έκανε η εταιρία Affimetrix, η οποία το 1996 παρήγαγε τις πρώτες μικροσυστοιχίες προς διάθεση στο εμπόριο [39]. Όσο το υλικό των μικροσυστοιχιών εξελίσσεται, δίνεται η δυνατότητα στις ερευνητικές ομάδες να αναπτύσσουν μικροσυστοιχίες με όλο και μεγαλύτερο αριθμό ανιχνευτών. Αυτό είχε ως αποτέλεσμα οι ερευνητές του Stanford να παρουσιάσουν την πρώτη μικροσυστοιχία ολόκληρου γονιδιώματος για τον ζυμομύκητα (μαγιά) [40]. Αντίστοιχα με την πάροδο των χρόνων όλο και περισσότερα βιολογικά πεδία έρευνας εύρισκαν εφαρμογή στις μικροσυστοιχίες. Έτσι η εταιρία Illumina το 2001 [41] δημιούργησε την πρώτη μικροσυστοιχία για τον εντοπισμό πολυμορφισμών στο DNA αναπτύσσοντας μία νέα τεχνολογία, τα λεγόμενα Beads, τα οποία είναι τρισδιάστατα πλέγμα από ανιχνευτές με σκοπό την επιτάχυνση της υβριδοποίησης, αλλά και την βελτίωση της αξιοπιστίας της. Η πρώτες κλινικές εφαρμογές των μικροσυστοιχιών εμφανίζονται το 2004, οι οποίες στόχευαν στην πρόβλεψη και την διάγνωση διαφόρων

νοσημάτων [42-44]. Σήμερα η εν λόγω τεχνολογία έχει καταφέρει να φτάσει σε επίπεδο διερεύνησης ολόκληρου του ανθρώπινου γονιδιώματος, όπου η επεξεργασία των παραγόμενων δεδομένων γίνεται με πολύ γοργούς ρυθμούς από εργαστήρια-επί-πλινθίου (Lab-on-Chip). Η ιστορική αναδρομή της τεχνολογίας με τα γεγονότα «ορόσημα» τα οποία αναφέρθηκαν ανωτέρω παρουσιάζονται στο Σχήμα 2.1.



Σχήμα 2.2: Η εξελικτική πορεία της τεχνολογίας των μικροσυστοιχιών.

Η ραγδαία εξέλιξη της τεχνολογίας των μικροσυστοιχιών μπορεί εύκολα να φανεί και από τον αριθμό των δημοσιεύσεων σε επιστημονικά περιοδικά και σε πρακτικά διεθνών συνεδρίων. Οι δημοσιεύσεις, οι οποίες σχετίζονται με την τεχνολογία των μικροσυστοιχιών και οι οποίες αφορούν ολόκληρο τον κύκλο ζωής των μικροσυστοιχιών, αυξάνονται με γεωμετρική πρόοδο τα τελευταία δέκα περίπου χρόνια. Αυτό συνιστά μια ένδειξη ότι η βιολογική επιστήμη έχει επενδύσει πολλά στην συγκεκριμένη τεχνολογία. Ο όγκος βέβαια τον δεδομένων που παράγονται απαιτεί ταυτόχρονη ενασχόληση και άλλων επιστημών με το συγκεκριμένο αντικείμενο, όπως της πληροφορικής. Το σχήμα 2.2 παρουσιάζει τον αριθμό των δημοσιεύσεων των σχετικών με την τεχνολογία μικροσυστοιχιών από την πρώτη εμφάνιση της μέχρι και σήμερα.



Σχήμα 2.3: Η ραγδαία εξέλιξη της τεχνολογίας μικροσυστοιχιών όπως αποτυπώνεται από τον αριθμό των δημοσιεύσεων τα τελευταία χρόνια.

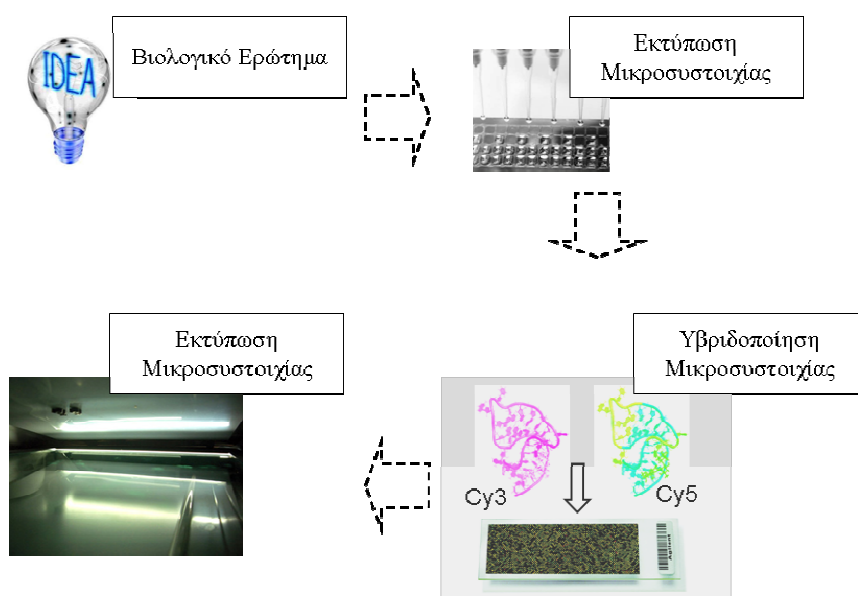
Από την ραγδαία εξέλιξη της τεχνολογίας απορρέουν πολλές παραλλαγές και ένας μεγάλος αριθμός εφαρμογών των μικροσυστοιχιών, με σκοπό την διερεύνηση των βιολογικών λειτουργιών.

2.2. Προετοιμασία του Πειράματος και βιολογικό πείραμα

Στο κεφάλαιο αυτό περιγράφονται όλα τα βήματα, τα οποία ακολουθούνται κατά την διάρκεια του κύκλου ζωής ενός πειράματος μικροσυστοιχιών, έως ότου καταλήξουμε στο ζητούμενο της παρούσας διδακτορικής διατριβής, το οποίο είναι η επεξεργασία των εικόνων των μικροσυστοιχιών. Παρά το μεγάλο πλήθος των εφαρμογών και την ποικιλία των ειδών των μικροσυστοιχιών, ένα πείραμα ακολουθεί μια τυποποιημένη διαδικασία. Στο τέλος του κεφαλαίου, θα περιγραφούν οι βάσεις δεδομένων με εικόνες μικροσυστοιχιών, οι οποίες χρησιμοποιήθηκαν για την αξιολόγηση των μεθόδων.

Όπως αναφέρθηκε στην εισαγωγή, η κατασκευή μιας μικροσυστοιχίας και τα στάδια του ίδιου του βιολογικού πειράματος είναι εκείνα τα οποία προηγούνται στον κύκλο ζωής. Το πέρας αυτών των σταδίων σηματοδοτεί την εξαγωγή της προς επεξεργασία εικόνας. Μέχρι την στιγμή της λήψης της εικόνας ακολουθούνται πέντε βασικά στάδια, όπως φαίνονται και στο σχήμα 2.3.

Αρχικά διευκρινίζεται ο λόγος για τον οποίο διεξάγεται το κάθε πείραμα. Ανάλογα με το βιολογικό ερώτημα που τίθεται επιλέγεται η εφαρμογή των μικροσυστοιχιών που πρέπει να χρησιμοποιηθεί, το είδος των αλληλουχιών που πρέπει να απομονωθούν, ποιοι δηλαδή συγκεκριμένα θα είναι οι ανιχνευτές που θα χρησιμοποιηθούν. Στην συνέχεια λαμβάνει χώρα η κατασκευή της μικροσυστοιχίας, κατά την οποία εκτυπώνονται οι ανιχνευτές επάνω στο υπόστρωμα. Ακολουθεί η ανάμειξη και η υβριδοποίηση των φθορίζοντων δειγμάτων με τους ανιχνευτές. Τέλος η μικροσυστοιχία σαρώνεται με σκοπό την εξαγωγή των εικόνων.

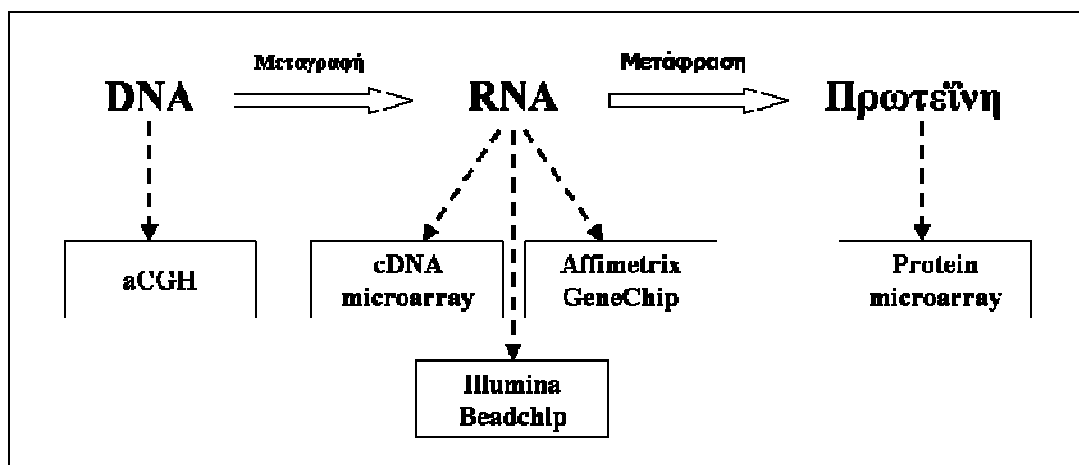


Σχήμα 2.4: Τα στάδια του κύκλου ζωής των μικροσυστοιχιών τα οποία προηγούνται της επεξεργασία των εξαγόμενων εικόνων.

2.2.1. Βιολογικό Ερώτημα

Από την πρώτη εμφάνιση των μικροσυστοιχιών, έχει αναπτυχθεί ένα ευρύ φάσμα εφαρμογών. Η πιο βασική από αυτές έχει να κάνει με κλινικές εφαρμογές, όπως για παράδειγμα η διάγνωση και η θεραπεία ασθενειών. Μέσω των μικροσυστοιχιών μπορεί να βρεθεί η προδιάθεση κάποιου ανθρώπου να εμφανίσει ένα νόσημα, αλλά και να προβλέψει επικίνδυνους παράγοντες κατά την διάρκεια της θεραπείας του. Πέραν των κλινικών εφαρμογών οι μικροσυστοιχίες έχουν εφαρμογές στην ανακάλυψη και ανάπτυξη φαρμάκων [45-47], αλλά και στην ανάλυση των τροφών [48-50].

Τα διαφορετικά είδη μικροσυστοιχιών μπορούν επίσης να κατηγοριοποιηθούν ανάλογα με της βιολογικές αλληλουχίες, οι οποίες χρησιμοποιούνται και οι οποίες υποβάλλονται σε υβριδοποίηση μαζί με τα προς εξέταση δείγματα. Στο σχήμα 2.4 παρουσιάζονται τα είδη των μικροσυστοιχιών που έχουν αναπτυχθεί μέχρι σήμερα, σε αντιστοιχία με το είδος των βιολογικών αλληλουχιών.



Σχήμα 2.5: Κατηγοριοποίηση των ειδών των μικροσυστοιχιών ανάλογα με της βιολογικές αλληλουχίες που χρησιμοποιούνται.

Όπως αναφέρθηκε και στην εισαγωγή η πρώτη εφαρμογή των μικροσυστοιχιών αφορά στην έκφραση των γονιδίων. Οι μικροσυστοιχιές αυτές έκαναν χρήση μονόκλωνων αλληλουχιών συμπληρωματικού DNA, οι οποίες υβριδοποιούνταν με αλληλουχίες RNA. Σκοπός τους είναι η ποσοτικοποίηση των επιπέδων της έκφρασης κάθε γονιδίου μέσω του ποσοστού της υβριδοποίησης των δειγμάτων με τους ανιχνευτές. Την ίδια εφαρμογή εξυπηρετεί και το Genechip της εταιρίας Affimetrix, το οποίο διατέθηκε λίγα χρόνια αργότερα στην αγορά. Τέλος αλληλουχίες από RNA κάνει χρήση και το Beadchip της εταιρίας Illumina. Η συγκεκριμένη όμως μικροσυστοιχία έχει δημιουργηθεί για τον εντοπισμό πολυμορφισμών στα δείγματα. Εκτός από τις μικροσυστοιχιές, η οποίες έχουν ως βάση τους το RNA, υπάρχουν και άλλες που έχουν ως βάση τους άλλες βιολογικές αλληλουχίες. Οι χρωμοσωματικές ανωμαλίες για παράδειγμα είναι ένα πεδίο που απαιτεί παραδοσιακά μεγάλη διερεύνηση, εξετάζοντας τον καρύοτυπο των χρωμοσωμάτων. Τα πρώτα χρόνια αυτό συνέβαινε με την χρήση μικροσκοπίου και με την μέθοδο FISH (Fluorescence in situ Hybridization). Αργότερα και για αυτήν την εφαρμογή αναπτύχθηκαν οι μικροσυστοιχιές aCGH (array *Comparative* Genomic Hybridization), οι οποίες υβριδοποιούν αλληλουχίες DNA [51]. Τέλος, παρά το γεγονός ότι οι DNA και οι RNA

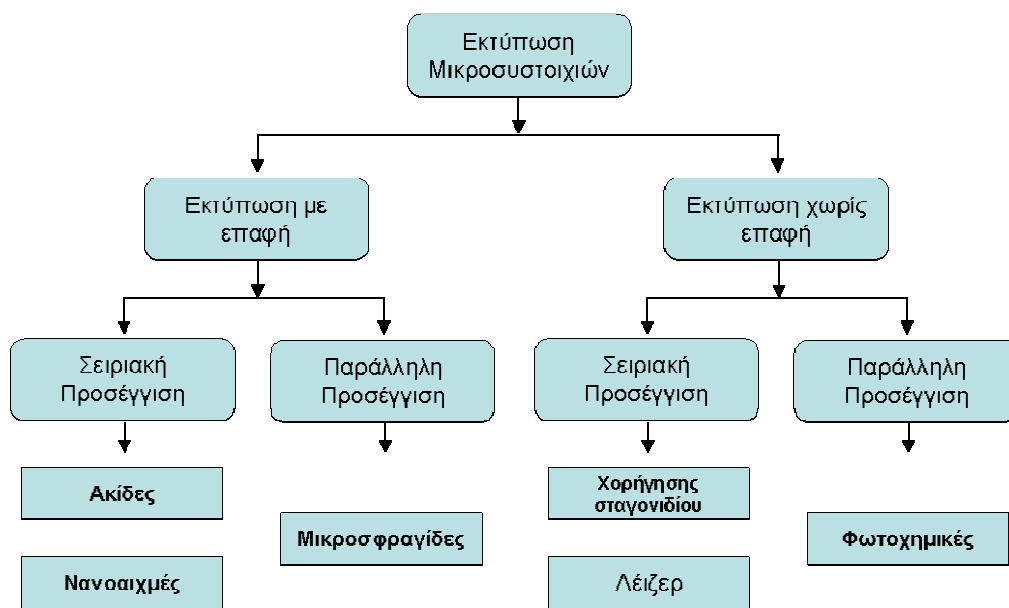
μικροσυστοιχίες μπορούν να ποσοτικοποιήσουν με ακρίβεια τα επίπεδα έκφρασης ενός γονιδίου σε ένα κύτταρο, δεν μπορούν να μετρήσουν το ποσοστό των πρωτεϊνών, οι οποίες μεταφράζονται σε αυτό. Το γεγονός αυτό οδήγησε στην δημιουργία ενός νέου είδους μικροσυστοιχιών των πρωτεϊνικών μικροσυστοιχιών [52].

2.2.2. Εκτύπωση των ανιχνευτών

Το δεύτερο στάδιο του κύκλου ζωής των μικροσυστοιχιών είναι η εκτύπωση των ανιχνευτών. Η διαδικασία της εκτύπωσης (printing) αφορά στην δημιουργία της μικροσυστοιχίας, και πιο συγκεκριμένα στον τρόπο με τον οποίο οι ανιχνευτές τοποθετούνται πάνω στο γυάλινο υπόστρωμα. Οι μέθοδοι για την εκτύπωση των μικροσυστοιχιών γενικά χωρίζονται σε δύο μεγάλες κατηγορίες, την εκτύπωση με επαφή (contact printing) και την εκτύπωση χωρίς επαφή (non-contact printing). Κατά την διάρκεια μιας εκτύπωσης με επαφή, η συσκευή εκτύπωσης έρχεται σε φυσική επαφή με την επιφάνεια του υποστρώματος, σε αντίθεση με την εκτύπωση χωρίς επαφή κατά την οποία καμιά φυσική επαφή δεν υπάρχει μεταξύ συσκευής εκτύπωσης και επιφανείας της μικροσυστοιχίας.

Οι δυο παραπάνω κατηγορίες κατηγοριοποιούνται περαιτέρω σε δύο υποκατηγορίες την σειριακή (serial) και την παράλληλη (parallel) προσέγγιση [53]. Η σειριακή προσέγγιση γίνεται με επαναλαμβανόμενες κινήσεις της συσκευής εκτύπωσης και για τον λόγο αυτό έχει περιορισμούς στην χρονική διάρκεια που απαιτείται για την ολοκλήρωση της διαδικασίας. Έτσι, για την εκτύπωση πολύ μεγάλου αριθμού ανιχνευτών χρησιμοποιείται η παράλληλη προσέγγιση, η οποία όμως δεδομένου ότι είναι πιο καινούργια και πιο απαιτητική τεχνολογικά είναι λιγότερο αξιόπιστη. Όλες βέβαια οι παραπάνω τεχνικές στοχεύουν στην αποδοτική, ομοιόμορφη, και ισόποση τοποθέτηση των ανιχνευτών επάνω στην επιφάνεια εκτύπωσης. Επίσης επιδιώκουν στο χαμηλό κόστος, αλλά και την αποφυγή προβλημάτων που μπορεί να προκύψουν, όπως της μόλυνσης της επιφάνειας, που όπως θα δούμε παρακάτω μπορεί να επηρεάσει σημαντικά την ανάλυση της εικόνας και κατ' επέκταση την επεξεργασία των εξαγόμενων δεδομένων. Στο σχήμα 2.5 παρουσιάζεται ένα δένδρογραμμα με τις υπάρχουσες τεχνικές για την εκτύπωση των μικροσυστοιχιών, τις οποίες θα δούμε συνοπτικά παρακάτω.

Όπως θα δούμε και παρακάτω πολύ σημαντικό στην ανάλυση της εικόνας των μικροσυστοιχιών και την αξιοπιστία των αλγορίθμων είναι η ομοιομορφία των κηλίδων στην εικόνα, και η ακρίβεια στις θέσεις που τοποθετούνται. Αυτό δεν μπορεί να επιτευχθεί πάντα στον βέλτιστο βαθμό διότι εξαρτάται από πολλούς παράγοντες όπως, οι ιδιότητες της επιφάνειας που θα τοποθετηθούν οι ανιχνευτές, και το ιξώδες των δειγμάτων. Δευτερεύοντες παράγοντες που μπορούν επίσης να επηρεάσουν την εκτύπωση είναι η ακρίβεια της ρομποτικής συσκευής, η υγρασία, η θερμοκρασία αλλά και η μόλυνση στον αέρα (π.χ. σκόνη) του περιβάλλοντος, στο οποίο λαμβάνει χώρα η εκτύπωση.



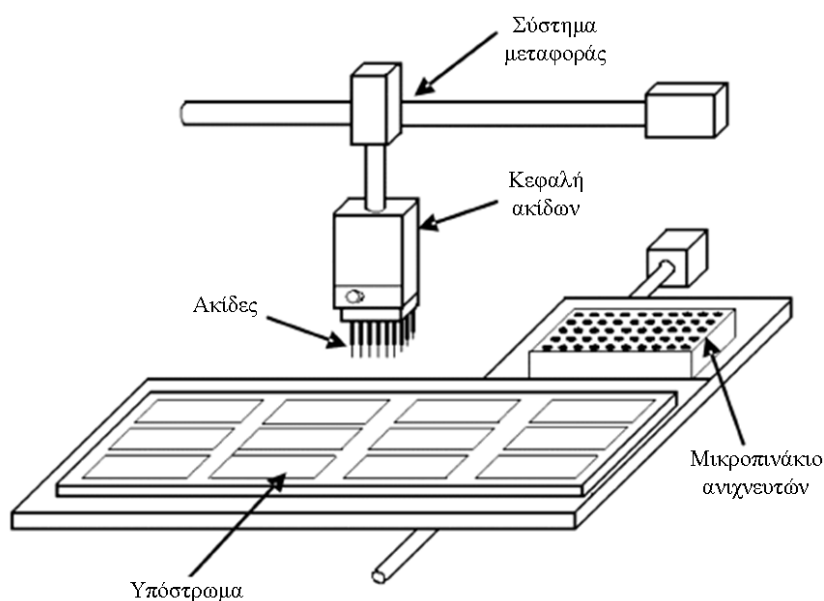
Σχήμα 2.6: Η τεχνικές για την εκτύπωση μικροσυστοιχιών.

Η μεγάλη ταχύτητα με την οποία γίνεται η εκτύπωση εισάγει μεγάλη αδράνεια στο δείγμα με αποτέλεσμα ένα μέρος του δείγματος να χύνεται έξω από τα πηγάδια της επιφάνειας που πρέπει να τοποθετηθεί. Σε αυτό το γεγονός μεγάλο ρόλο διαδραματίζει και η τάση επιφανείας του δείγματος. Όλα τα παραπάνω αναφέρονται γιατί όπως θα δούμε δημιουργούν διάφορα τεχνουργήματα (artefacts) στην εικόνα κατά την διάρκεια της σάρωσης.

2.2.2.1. Εκτύπωση με επαφή

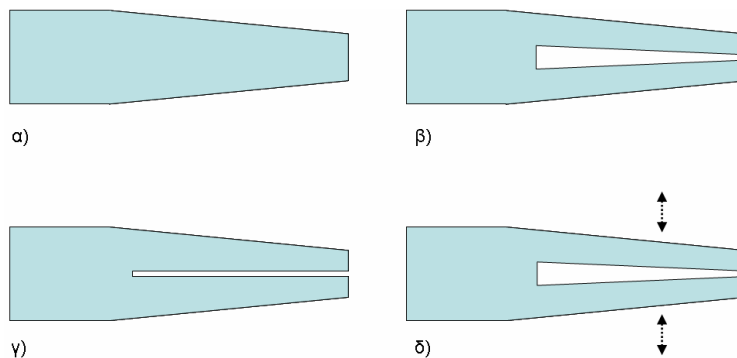
Όπως αναφέρθηκε και ανωτέρω η εκτύπωση με επαφή αφορά στις τεχνικές, στις οποίες η συσκευή εκτύπωσης και η μικροσυστοιχία έχουν φυσική επαφή κατά την διάρκεια της εκτύπωσης. Όπως φαίνεται και στο σχήμα 2.5 για την κατηγορία αυτή χρησιμοποιούνται εκτυπωτές ακίδων (είτε συμπαγών ακίδων (solid pins), είτε ακίδων με σχισμή (split pins)) ή εκτυπωτές νανο-αιχμών (nano-tips), για σειριακή εκτύπωση. Αντίθετα για παράλληλη εκτύπωση χρησιμοποιείτε η τεχνική των μικροσφραγίδων (microstamps).

Μια από τις περισσότερο διαδεδομένες τεχνικές με επαφή για την εκτύπωση των μικροσυστοιχιών είναι η εκτύπωση με ακίδες [54]. Μια ρομποτική συσκευή χρησιμοποιεί μια σειρά από ακίδες, οι οποίες εκτυπώνουν στην επιφάνεια τους ανιχνευτές. Μια τέτοια συσκευή φαίνεται στο σχήμα 2.6.



Σχήμα 2.7: Συσκευή εκτύπωσης μικροσυστοιχιών με ακίδες.

Η πρώτες μικροσυστοιχίες που κατασκευάστηκαν στο Πανεπιστήμιο του Στάνφορντ δημιουργήθηκαν χρησιμοποιώντας ακίδες [38]. Η τεχνική της εκτύπωσης με ακίδες συνέχισε να χρησιμοποιείται πάρα πολύ κυρίως σε μη εμπορικό επίπεδο. Ένας τομέας έρευνας για την εκτύπωση με ακίδες αφορά τον σχεδιασμό των ίδιων των ακίδων. Στο σχήμα 2.7 παρουσιάζονται τέσσερα είδη ακίδων που έχουν χρησιμοποιηθεί για την εκτύπωση μικροσυστοιχιών [55-56].



Σχήμα 2.8: Διαφορετικοί σχεδιασμοί ακίδων: α) συμπαγής ακίδα β), γ), και δ) διαφορετικές ακίδες με σχισμή.

Ο πιο απλός σχεδιασμός για μια ακίδα είναι η συμπαγής ακίδα, όπως φαίνεται στο σχήμα 2.7α. Ο περισσότερο διαδεδομένος σχεδιασμός ακίδων είναι οι ακίδες με σχισμή. Το μεγάλο πλεονέκτημα αυτού του σχεδιασμού έγκειται στο γεγονός ότι μέσα στην σχισμή μπορεί να χωρέσει μεγαλύτερη ποσότητα του ανιχνευτή. Με αυτόν τον τρόπο ο εκτυπωτής, εκτυπώνει σειριακά αρκετούς ανιχνευτές χωρίς να χρειάζεται να λαμβάνει τον ανιχνευτή κάθε φορά. Αντίθετα οι εκτυπωτές με συμπαγείς ακίδες κάνουν διαρκώς την κίνηση από το πηγάδι φόρτωσης του ανιχνευτή, στο υπόστρωμα εκτύπωσης. Τέλος, όπως φαίνεται στα σχήματα 2.7 β), γ) και δ) υπάρχουν διαφορετικοί τύποι σχισμών [57].

Η εκτύπωση με ακίδες είναι μια σειριακή τεχνική κατά την οποία είτε μια ακίδα είτε μια ομάδα από ακίδες εκτυπώνει επαναληπτικά επάνω στο υπόστρωμα τους ανιχνευτές. Μια εναλλακτική τεχνική εκτύπωσης είναι η εκτύπωση με μικροσφραγίδες [58-59]. Με τις μικροσφραγίδες χιλιάδες ανιχνευτές εκτυπώνονται ταυτόχρονα, σε μια διαδικασία, η οποία είναι εύκολα υλοποιήσιμη. Οι ανιχνευτές προσροφώνται σε μια επιφάνεια με συγκεκριμένο μοτίβο (patterned surface), η οποία θυμίζει σφραγίδα, και στην συνέχεια μεταφέρονται επάνω στο υπόστρωμα με φυσική επαφή. Πολύ σημαντικό ρόλο σε αυτήν την διαδικασία διαδραματίζει το υλικό, από το οποίο είναι κατασκευασμένες οι μικροσφραγίδες. Συνήθως, για να επιτευχθεί καλύτερη επαφή της σφραγίδας και του υποστρώματος γίνεται χρήση ελαστομερών υλικών.

Η τελευταία τεχνική εκτύπωσης μικροσυστοιχιών με επαφή, είναι η τεχνική των νανοαιχμών [60-61]. Η τεχνική αυτή αναπτύχθηκε λόγω των αυξημένων απαιτήσεων της τεχνολογίας των μικροσυστοιχιών να τοποθετήσει περισσότερους ανιχνευτές επάνω σε μια μικροσυστοιχία. Έτσι με την τεχνική αυτή εκτυπώνονται ανιχνευτές σε κλίμακα μικρότερη της τάξεως μεγέθους των μικρόμετρων. Η βάση της τεχνικής αυτής είναι η Μικροσκοπία Ατομικής Δύναμης (Atomic Force Microscopy – AFM). Σύμφωνα με αυτήν κατά την επαφή μιας AFM αιχμής με το υπόστρωμα, το δείγμα τοποθετείται επάνω στο υπόστρωμα λόγω τριχοειδούς δράσεως.

2.2.2.2.Εκτύπωση Χωρίς επαφή

Σε αντίθεση με τις τεχνικές με επαφή, οι τεχνικές χωρίς επαφή δεν επιτρέπουν την άμεση επαφή του ρομποτικού εκτυπωτή με το υπόστρωμα. Οι τεχνικές αυτές είναι βασισμένες σε φωτοχημικές μεθόδους και σε ακτίνες λέιζερ. Το βασικό πλεονεκτήματα των τεχνικών χωρίς επαφή είναι η ελαχιστοποίηση του θορύβου στις παραγόμενες εικόνες. Αυτό επιτυγχάνεται για δύο λόγους. Το γεγονός ότι κανένα σημείο του εκτυπωτή δεν έρχεται σε επαφή με το υπόστρωμα, αλλά διατηρείται μεταξύ τους απόσταση, κρατάει το υπόστρωμα καθαρό από διαφόρους τύπους μολύνσεων. Αυτό απαλλάσσει τις εικόνες από τεχνουργήματα λόγω μολύνσεων. Εκτός από την μόλυνση του υποστρώματος όμως, με την χρήση τεχνικών χωρίς επαφή αποφεύγετε η υπερχειλίση δείγματος σε άλλα σημεία του υποστρώματος. Για τους παραπάνω λόγους οι εικόνες οι οποίες παράγονται χωρίς την επαφή του εκτυπωτή με το υπόστρωμα παρέχουν εικόνες πολύ καλύτερης ποιότητας.

Οι φωτοχημικές τεχνικές εκτύπωσης είναι βασισμένες στην χημική επεξεργασία του υποστρώματος και την έκθεση του στην UV ακτινοβολία. Οι βασικές προσεγγίσεις αυτής της κατηγορίας είναι η εκτύπωση με φωτολιθογραφία (photolithography) [62-65], και η φωτοχημική προτυποποίηση (photochemical patterning) [62,66-67]. Στις τεχνικές φωτολιθογραφίας ένα φωτοαντιδραστικό υλικό επιστρώνεται επάνω στο υπόστρωμα, και κατόπιν εκτίθεται σε ακτινοβολία UV μέσω μιας φωτομάσκας. Με τον τρόπο αυτό δημιουργούνται επάνω στο υπόστρωμα ανοιχτές περιοχές, στις οποίες θα προσκολληθούν οι ανιχνευτές. Στην συνέχεια το υπόστρωμα βυθίζεται σε έναν διαλύτη για να καθαριστεί από το φωτοαντιδραστικό υλικό. Οι τεχνικές της φωτοχημικής προτυποποίησης δεν

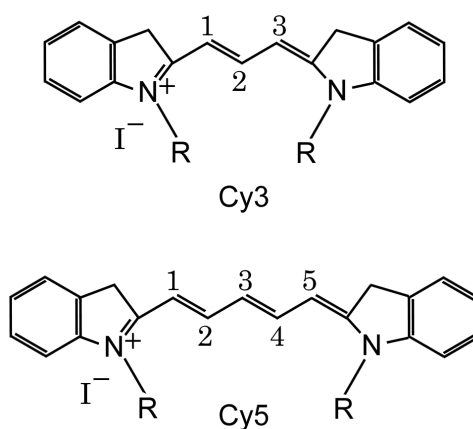
διαφέρουν πολύ από αυτές της φωτολιθογραφίας. Η μόνη διαφορά είναι ότι δεν επιστρώνεται το υπόστρωμα από το φωτοπαθές υλικό, αλλά χρησιμοποιούνται μόρια φωτοχημικών ενώσεων, τα οποία είτε ενεργοποιούνται είτε απενεργοποιούνται από την έκθεση τους στην ακτινοβολία UV.

Μια ευρέως διαδεδομένη τεχνική εκτύπωση χωρίς επαφή είναι η λεγόμενη τεχνική χορήγησης σταγονιδίων (Droplet Dispensing). Η κατηγορία αυτή όπως φανερώνει και η ονομασία της στοχεύει στην τοποθέτηση σταγονιδίων των ανιχνευτών στο υπόστρωμα. Ο Zeng [68] πρότεινε την χρησιμοποίηση των ήδη υπαρχόντων τεχνικών με ακίδες, χωρίς όμως αυτές να έχουν επαφή με το υπόστρωμα. Έτσι η τεχνική αυτή χρησιμοποιεί τις ακίδες με σχισμή, οι οποίες χρησιμοποιούνται και στην αντίστοιχη τεχνική με επαφή. Αφού η ακίδα φορτωθεί με τον ανιχνευτή επιταχύνεται προς το υπόστρωμα και επιβραδύνεται απότομα πριν την επαφή με το υπόστρωμα, με αποτέλεσμα ο ανιχνευτής να τοποθετείται επάνω στο υπόστρωμα λόγω ορμής. Στην ίδια κατηγορία ανήκει και η προσέγγιση η οποία εκμεταλλεύεται την τεχνολογία της εκτύπωσης με έκχυση μελάνης (ink-jet printing) [69]. Η προσέγγιση αυτή μάλιστα έχει πολύ μικρότερο κόστος από την προηγούμενη, η οποία απαιτεί ρομποτικό εκτυπωτή μεγάλης ακρίβειας.

Η πιο πρόσφατη τεχνική για την εκτύπωση των ανιχνευτών είναι η εκτύπωση με λέιζερ, η οποία έχει χρησιμοποιηθεί μέχρι στιγμής για την εκτύπωση πρωτεϊνικών μικροσυστοιχιών [70]. Κατά την εκτύπωση χρησιμοποιείται ένας δίσκος χαλαζία, επάνω στον οποίο επιστρώνεται μείγμα βιολογικών δειγμάτων γλυκερίνης [71-72]. Ένα παλμικό λέιζερ σαρώνει την επιφάνεια του δίσκου εξατμίζοντας μικρές περιοχές της επίστρωσης. Κατά την διαδικασία απελευθερώνονται σταγονίδια, τα οποία συσσωρεύονται στο υπόστρωμα. Τα σταγονίδια είναι πολύ μικρότερα από αυτά που τοποθετούνται στο υπόστρωμα με τις τεχνικές των ακίδων. Το γεγονός αυτό καθιστά την συγκεκριμένη τεχνική εκτύπωση πλεονεκτικότερη επειδή απαιτούνται πολύ μικρότερα ποσά βιολογικών δειγμάτων, σε σχέση με τις άλλες τεχνικές, αλλά και επειδή επιτυγχάνονται πολύ μικρότερες κηλίδες [72].

2.2.3. Υβριδοποίηση των δειγμάτων με τους ανιχνευτές

Η βασική ιδέα σε ένα πείραμα μικροσυστοιχιών είναι ο προσδιορισμός του ποσοστού της υβριδοποίησης των ανιχνευτών με τα δείγματα. Ανάλογα με την βιολογική υπόθεση, έχουν επιλεγεί οι κατάλληλοι ανιχνευτές, οι οποίοι είναι εκτυπωμένοι επάνω στην μικροσυστοιχία από κατασκευής, όπως είδαμε στα προηγούμενα εδάφια. Στην συνέχεια λαμβάνονται και απομονώνονται τα προς εξέταση δείγματα. Ένα σημαντικό σημείο για την διαδικασία της υβριδοποίησης είναι η διάκριση των δειγμάτων. Τα δύο διαφορετικά δείγματα (π.χ. κανονικό και παθογενές) πρέπει να είναι διακριτά μετά το πείραμα ώστε να συμπεράνουμε ποιο από τα δύο έχει υβριδοποιηθεί και ποιο όχι. Για τον λόγο αυτό, προτού λάβει χώρα η υβριδοποίηση των δειγμάτων με τους ανιχνευτές της μικροσυστοιχίας, στα προς εξέταση δείγματα προσκολλούνται φθορίζουσες ουσίες. Οι ουσίες που έχουν χρησιμοποιηθεί περισσότερο στην τεχνολογία των μικροσυστοιχιών είναι οι Cy3 και Cy5. Τις φθορίζουσες ουσίες Cy3 και Cy5 εισήγαγε πρώτος ο Ernst [73] το 1989. Μια τυπική μορφή των δύο ουσιών φαίνεται στο σχήμα 2.8.



Σχήμα 2.9: Οι δομή των φθορίζουσών ουσιών Cy3 και Cy5.

Οι ουσία Cy3 διεγείρεται κατά μέγιστο σε μήκος κύματος 550 nm ενώ εκπέμπει κατά μέγιστο σε μήκος κύματος 570 nm, δηλαδή στο παράθυρο του ορατού πράσινου του ηλεκτρομαγνητικού φάσματος. Αντίθετα η ουσία Cy5 διεγείρεται κατά μέγιστο σε μήκος κύματος 649 nm ενώ εκπέμπει κατά μέγιστο σε μήκος κύματος 670 nm, δηλαδή στο παράθυρο του ορατού κόκκινου του ηλεκτρομαγνητικού φάσματος. Πέραν των ουσιών Cy3 και Cy5 στην βιβλιογραφία έχουν χρησιμοποιηθεί και άλλες φθορίζουσες ουσίες.

Μερικές από αυτές παρουσιάζονται στον πίνακα 2.1 μαζί με τα μέγιστα μήκη κύματος στα οποία διεγείρονται και εκπέμπουν.

Μετά την προσκόλληση των φθορίζουσών ουσιών τα δύο δείγματα αναμιγνύονται με νερό και τοποθετούνται στην μικροσυστοιχία σε όλες τις θέσεις όπου υπάρχουν ανιχνευτές. Σύμφωνα με τις αρχές της υβριδοποίησης που εξετάσαμε στο προηγούμενο κεφάλαιο, οι συμπληρωματικές αλυσίδες δειγμάτων και ανιχνευτών θα δημιουργήσουν υβρίδια. Μετά από λίγες ώρες, η διαφάνεια πλένεται ώστε να απομακρυνθούν τα δείγματα, τα οποία δεν έχουν υβριδοποιηθεί. Στο σημείο αυτό το βιολογικό πείραμα έχει ολοκληρωθεί και οι μικροσυστοιχία είναι έτοιμη για να σαρωθεί ,και να εξαχθούν οι εικόνες.

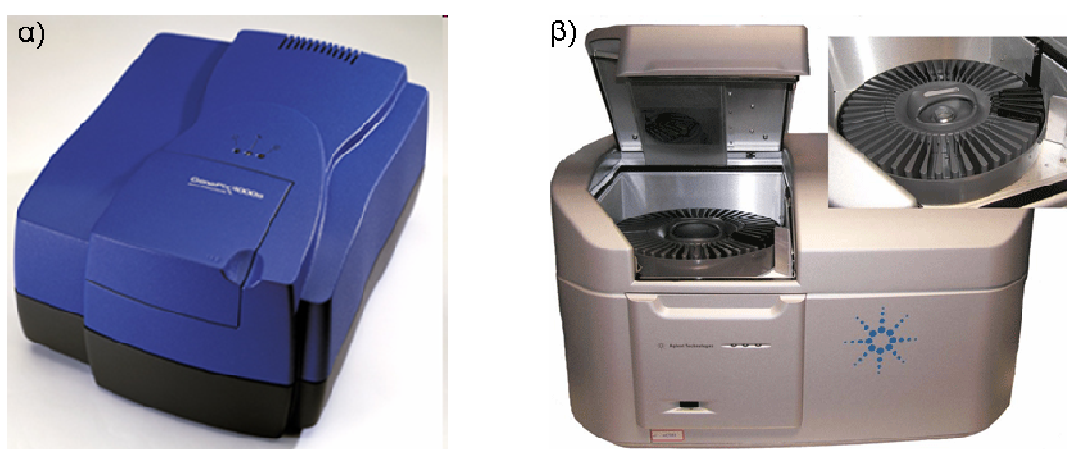
Πίνακας 2.1: Διαφορετικές φθορίζουσες ουσίες που χρησιμοποιούνται στην τεχνολογία μικροσυστοιχιών. Τα μέγιστα μήκη κύματος διέγερσης και εκπομπής φαίνονται στην δεύτερη και την Τρίτη στήλη του πίνακα αντίστοιχα.

| Dye | Absorption maximum [nm] | Emission maximum [nm] |
|----------------|----------------------------|--------------------------|
| FITC | 490 | 520 |
| Alexa 488 | 495 | 519 |
| Alexa 568 | 578 | 603 |
| JOE | 522 | 550 |
| Cy2 | 489 | 506 |
| Cy3 | 550 | 570 |
| Cy5 | 649 | 670 |
| Cy7 | 743 | 767 |
| BODIPY 630/650 | 625 | 640 |
| BODIPY 650/665 | 646 | 660 |

2.2.4. Σάρωση

Αφού τα δείγματα υβριδοποιηθούν η μικροσυστοιχία τοποθετείται σε σαρωτή (scanner). Η σάρωση της μικροσυστοιχίας γίνεται και στα δύο μήκη κύματος, στα οποία εκπέμπουν οι φθορίζουσες ουσίες. Για της φθορίζουσες ουσίες Cy3 και Cy5 για παράδειγμα, οι σαρωτές

συνήθως χρησιμοποιούν μήκη κύματος εκπομπής του λέιζερ 532 nm και 635 nm ενώ ταυτόχρονα χρησιμοποιούν ζωνοπερατά φίλτρα σε μήκη κύματος 550-600 nm και 655-695 nm με σκοπό την αποφυγή του θορύβου υποβάθρου. Με τον τρόπο αυτόν καθίσταται πιο εύκολος ο διαχωρισμός των δύο δειγμάτων, τα οποία εκπέμπουν σε διαφορετικά μήκη κύματος. Σκοπός της διαδικασίας της σάρωσης είναι βέβαια η ποσοτικοποίηση της φωτεινότητας που εκπέμπουν τα δείγματα στην θέση του κάθε ανιχνευτή. Στο παρακάτω σχήμα 2.9 παρουσιάζονται δύο ευρέως χρησιμοποιούμενοι σαρωτές μικροσυστοιχιών, των εταιριών Genepix και Agilent.

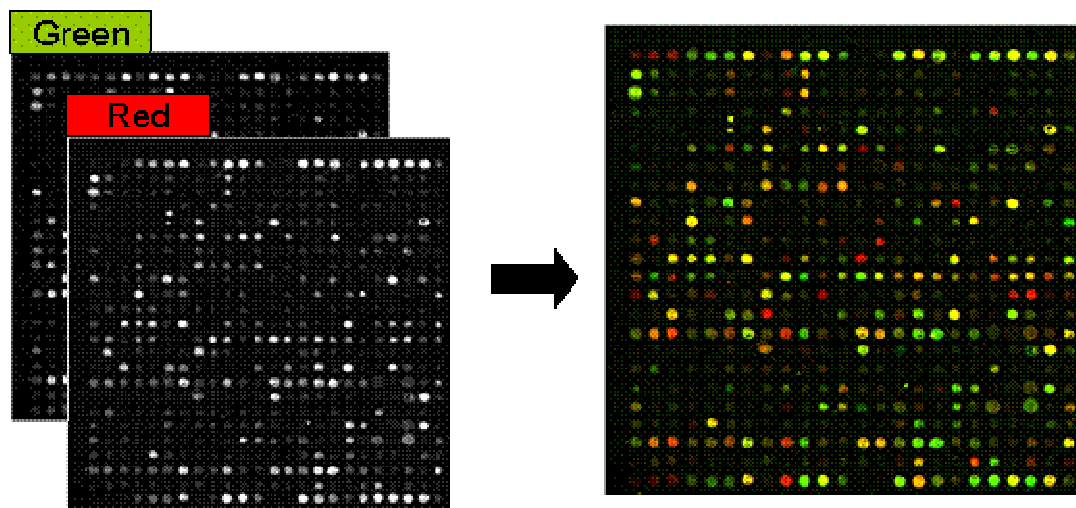


Σχήμα 2.10: Δύο πολύ γνωστοί σαρωτές μικροσυστοιχιών α) της εταιρίας Genepix και β) της εταιρίας Agilent.

Από την σάρωση προκύπτουν δύο εικόνες επιπέδων γκρι 16-bit ή 8-bit, στις οποίες αποτυπώνονται οι ανιχνευτές που έχουν υβριδοποιηθεί με τα δύο δείγματα, δεδομένου ότι φανερώνουν σε ποια σημεία της εικόνας έχουν κατακρατηθεί οι φθορίζουσες ουσίες. Η δύο εικόνες επιπέδων γκρι θεωρούνται η κάθε μία ως ένα κανάλι και ψευδοχρωματίζονται με κόκκινο και πράσινο χρώμα αντίστοιχα. Κατ' αυτόν τον τρόπο προκύπτει η δικαναλική εικόνα μικροσυστοιχιών, η οποία είναι πολύ βολική στην οπτική αναπαράσταση του πειράματος. Η εικόνα αυτή είναι στην πραγματικότητα μια έγχρωμη RGB εικόνα της οποίας το μπλε κανάλι είναι μηδενικό. Παρατηρώντας μια έγχρωμη εικόνα μικροσυστοιχιών, μπορούν να εξαχθούν άμεσα τα ακόλουθα τέσσερα συμπεράσματα ανάλογα με το χρώμα της εικόνας στη περιοχή του κάθε ανιχνευτή: α) εάν είναι πράσινο τότε ο ανιχνευτής έχει υβριδοποιηθεί μόνο με το δείγμα το οποίο ψευδοχρωματίστηκε ως πράσινο, β) αν είναι κόκκινο τότε, αντίστοιχα ο ανιχνευτής έχει υβριδοποιηθεί μόνο με το

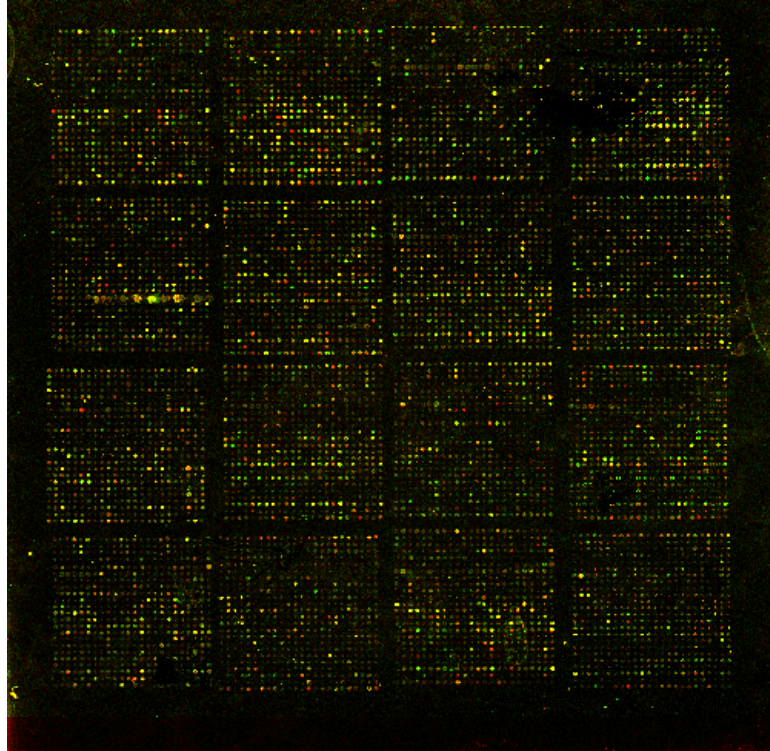
δείγμα που ψευδοχρωματίστηκε κόκκινο, γ) αν είναι κίτρινο τότε έχουν υβριδοποιηθεί και τα δύο δείγματα με τον ανιχνευτή και τέλος, δ) αν είναι μαύρο τότε δεν έχει υβριδοποιηθεί κανένα από τα δύο δείγματα με τον ανιχνευτή.

Πέραν του γεγονότος ότι με την έγχρωμη εικόνα μικροσυστοιχιών γίνεται εύκολα αντιληπτό ποιο από τα δείγματα έχει υβριδοποιηθεί με ποιους ανιχνευτές, η εικόνα αυτή δίνει και την δυνατότητα εφαρμογής πολυκαναλικών τεχνικών επεξεργασίας εικόνας. Στο σχήμα 2.10 φαίνεται ένα τμήμα μίας εικόνας μικροσυστοιχιών όπως αυτή προκύπτει κατά την σάρωση. Αριστερά στο σχήμα παρουσιάζονται οι δύο εικόνες επιπέδων γκρι, ενώ δεξιά φαίνεται η δικαναλική εικόνα.



Σχήμα 2.11: Τα δύο κανάλια στο αριστερό μέρος της εικόνας και οι έγχρωμη εικόνα μικροσυστοιχιών.

Χρησιμοποιώντας την ορολογία της επεξεργασίας των εικόνων των μικροσυστοιχιών, η κάθε κυκλική περιοχή που προέρχεται από την υβριδοποίηση ενός ανιχνευτή με τα δείγματα ονομάζεται κηλίδα (spot). Λόγου του μεγάλου αριθμού των ανιχνευτών που υπάρχουν σε μια μικροσυστοιχία, και κατ' επέκταση του μεγάλου αριθμού κηλίδων σε μια εικόνα μικροσυστοιχιών, συνηθίζεται η χωροθέτηση των ανιχνευτών σε τετράγωνες περιοχές πάνω στην μικροσυστοιχία. Αυτές οι περιοχές ονομάζονται συστάδες ή υποπλέγματα (blocks or subgrids).



Σχήμα 2.12: Εικόνα μικροσυστοιχιών με 16 συστάδες και 576 κηλίδες σε κάθε συστάδα

Η εικόνα που παρουσιάζεται στο σχήμα 2.11 είναι μια εικόνα από την βάση δεδομένων του Πανεπιστημίου του Στάνφορντ, η οποία όπως φαίνεται περιέχει 16 συστάδες από κηλίδες (4 στον κατακόρυφο και 4 στο οριζόντιο άξονα). Η κάθε συστάδα, περιέχει 24 κηλίδες στον κατακόρυφο άξονα, και 24 στον οριζόντιο. Δηλαδή η κάθε συστάδα αυτής της εικόνας περιέχει 576 κηλίδες, ενώ συνολικά ολόκληρη η εικόνα περιέχει 9216 κηλίδες. Γενικά, όπως εύκολα γίνεται αντιληπτό ο αριθμός των κηλίδων και των συστάδων είναι καθαρά θέμα κατασκευής της μικροσυστοιχίας κατά την διαδικασία της εκτύπωσης, με αποτέλεσμα να υπάρχει μεγάλη ποικιλομορφία των εικόνων σε αυτό το επίπεδο.

2.3.Σύνολα Δεδομένων

Κατά την διεξαγωγή της παρούσας διατριβής χρησιμοποιήθηκαν πραγματικές αλλά προσομοιωμένες εικόνες μικροσυστοιχιών με σκοπό την αξιολόγηση των μεθόδων οι οποίες αναπτύχθηκαν. Οι πραγματικές εικόνες προέρχονται κυρίως από την βάση δεδομένων του Στάνφορντ, στην οποία υπάρχει πρόσβαση μέσω διαδικτύου. Επίσης χρησιμοποιήθηκαν σαρωμένες εικόνες, οι οποίες λήφθηκαν στα πλαίσια ερευνητικού έργου χρηματοδοτούμενου από την Ευρωπαϊκή Ένωση και με το όποιο η παρούσα

διατριβή είχε άμεση συνεργασία. Οι προσομοιωμένες εικόνες δημιουργήθηκαν από τον πιο διαδεδομένο προσομοιωτή μικροσυστοιχιών. Όπως θα δούμε παρακάτω, διενεργήθηκαν διάφορες παρεμβάσεις στον κώδικά του προσομοιωτή ώστε να παρέχει επισημείωση των εικόνων με πληροφορία ανά εικονοστοιχείο, καθώς επίσης και για να παρέχει εικόνες στις οποίες οι κηλίδες είναι χωροθετημένες σε εξαγωνικό πλέγμα.

2.3.1. Βάση δεδομένων του Στάνφορντ

Η βάση δεδομένων του Στάνφορντ [74] είναι μια από της μεγαλύτερες βάσεις δεδομένων μικροσυστοιχιών. Περιέχει έναν πολύ μεγάλο αριθμό από πειράματα μικροσυστοιχιών, για κάθε ένα από τα οποία παρέχει ανεπεξέργαστες εικόνες, όπως αυτές παράγονται από την διαδικασία της σάρωσης, αλλά και κανονικοποιημένα δεδομένα όπως αυτά εξάγονται από τις εικόνες με την βοήθεια γνωστών λογισμικών πακέτων (κυρίως με χρήση του πακέτου το ScanAlyze [75]). Η συγκεκριμένη βάση παρέχει επίσης την δυνατότητα ανάλυσης και απεικόνισης των δεδομένων με την χρήση διαδικτυακών διεπαφών. Οι δύο βασικοί λόγοι για τους οποίους δημιουργήθηκε η βάση δεδομένων του Στάνφορντ ήταν η εξυπηρέτηση των ερευνητών του ίδιου του Πανεπιστημίου του Στάνφορντ, αλλά και γενικά η δημοσιοποίηση των δεδομένων αυτών. Για τους παραπάνω λόγους η βάση δεδομένων του Στάνφορντ επιτρέπει στους χρήστες την φόρτωση των πειραμάτων τους στην βάση μέσω διαδικτύου, χρησιμοποιώντας συγκεκριμένο πρωτόκολλο για την μορφή των δεδομένων και των εικόνων. Για παράδειγμα, η βάση δέχεται δεδομένα τα οποία έχουν προκύψει από τα λογισμικά πακέτα GenePix [76] και ScanAlyze. Συγκεκριμένα όσον αφορά τις εικόνες, η βάση δεδομένων του Στάνφορντ παρέχει, τις TIFF εικόνες επιπέδων του γκρι (16-bit) και την δικαναλική εικόνα GIF με την οποία επιτρέπεται στον χρήστη να βλέπει μια απεικόνιση των δεδομένων. Από την εικόνα αυτή ο χρήστης μπορεί να λάβει τις πληροφορίες για τον ανιχνευτή που βρίσκεται σε κάθε μια κηλίδα. Επίσης μπορεί να λάβει τα δεδομένα που έχουν προκύψει για την υβριδοποίηση του ανιχνευτή αυτού με τα δείγματα, από την επεξεργασία με τα πακέτα που προαναφέρθηκαν.

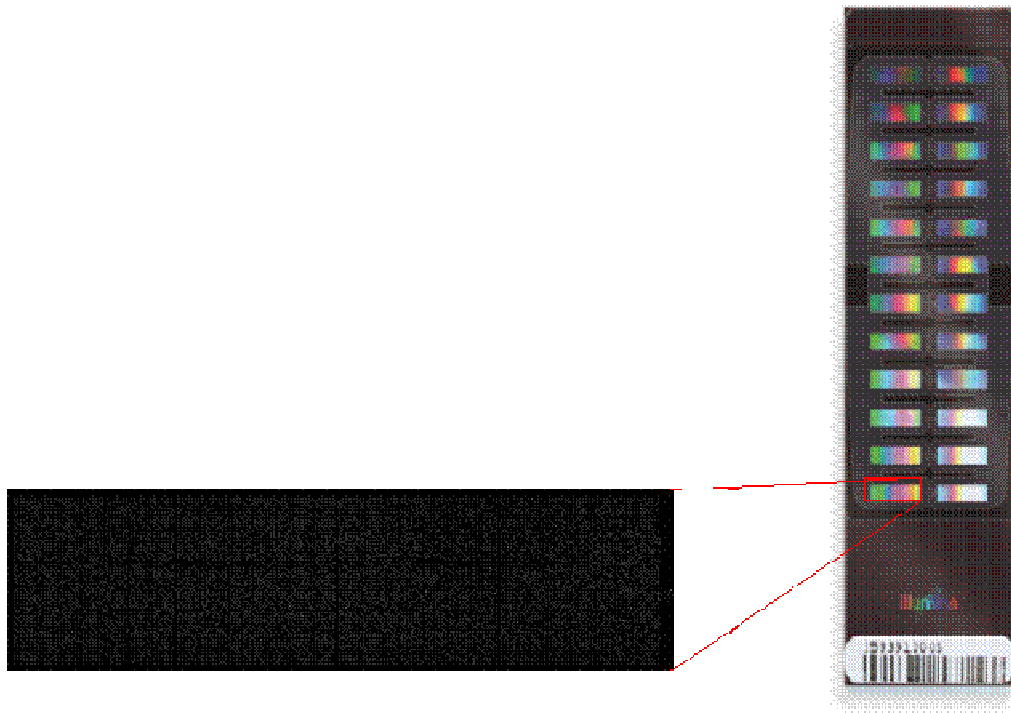
2.3.2. Εικόνες από την μικροσυστοιχία CNV370 της Illumina

Ένα άλλο σύνολο εικόνων που χρησιμοποιήθηκε για την αξιολόγηση των μεθόδων έχει εξαχθεί από πειράματα σε σταθμό της εταιρίας Illumina (Illumina Station) [77]. Η παραγωγή των εικόνων έγινε στα πλαίσια χρηματοδοτούμενου προγράμματος από την ευρωπαϊκή ένωση, σχετικό με την έγκαιρη διάγνωση της ρευματοειδούς αρθρίτιδας και της σκλήρυνσης κατά πλάκας. Συγκεκριμένα, οι εικόνες προέρχονται από την μικροσυστοιχία CNV370, η οποία κάνει χρήση της τεχνολογίας των beads που έχει αναπτύξει η ίδια εταιρία. Η τεχνολογία αυτή αφορά στην ταχύτερη και πιο αξιόπιστη υβριδοποίηση των δειγμάτων με τους ανιχνευτές. Για τον σκοπό αυτό για κάθε έναν από τους ανιχνευτές δημιουργείται έναν 3-διάστατο πλέγμα από την αλληλουχία του.

Οι 3 βασικές ιδιότητες της μικροσυστοιχίας συνοψίζονται παρακάτω:

- Η ανάλυση της παραλλαγής αριθμού αντιγράφων (Copy Number Variation) εξετάζει τις παραλλαγές του ανθρώπινου γονιδιώματος. Οι παραλλαγές αυτές, καθώς και η συσχέτιση τους με διάφορες ασθένειες βρίσκονται στο επίκεντρο της έρευνας των βιολόγων. Η μικροσυστοιχία HumanCNV370-Duo BeadChip δημιουργήθηκε στα πλαίσια του παγκόσμιου προγράμματος HumanHar, και αποτελεί την βασική μικροσυστοιχία για τις μελέτες συσχέτισης ασθενειών με το ανθρώπινο γονιδίωμα.
- Η μικροσυστοιχία αναλύει δύο δείγματα ταυτόχρονα, υβριδοποιώντας τα με πάνω από 370,000 ανιχνευτές, κάνοντας χρήση της ανάλυσης Infinium χωρίς να χρειάζεται την αλυσιδωτή αντίδραση πολυμεράσης.
- Η μικροσυστοιχία περιέχει περίπου 4.3 μεγαβάσεις από συγκεκριμένη περιοχή του ανθρώπινου γονιδιώματος, η οποία περιέχει πολλά γονίδια τα οποία έχουν ήδη συσχετιστεί με διάφορα αυτοάνοσα νοσήματα όπως ο διαβήτης τύπου I, η σκλήρυνση κατά πλάκας, η μαλάρια κ.α..

Το σχήμα 2.12 δείχνει την μικροσυστοιχία CNV370 και την εικόνα, η οποία εξάγεται από ένα μέρος της μικροσυστοιχίας αυτής.



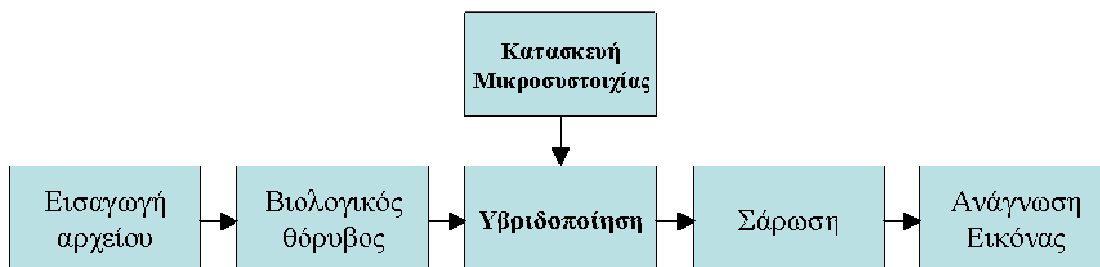
Σχήμα 2.13: Η μικροσυστοιχία Beadchip CNV370 της εταιρίας Illumina και η εξαγόμενη εικόνα.

Η βασική ιδιότητα των εικόνων, που προέρχονται από την μικροσυστοιχία CNV370 και η οποία επηρέασε την εργασία που έγινε στα πλαίσια της διατριβής, σχετίζεται με τη δομή που έχουν η εικόνες όσον αφορά στην χωροθέτηση των ανιχνευτών. Σε αντιπαράθεση με τις συνηθισμένες εικόνες μικροσυστοιχιών, στις οποίες η κηλίδες βρίσκονται τοποθετημένες σε γραμμές και στήλες, στην περίπτωση της συγκεκριμένης μικροσυστοιχίας οι κηλίδες είναι τοποθετημένες στις κορυφές εξαγωνικού πλέγματος. Το γεγονός αυτό, όπως θα εξετάσουμε και στο τέταρτο κεφάλαιο δημιουργεί αυξημένες απαιτήσεις από έναν αλγόριθμο εντοπισμού των κηλίδων.

2.3.3. Προσομοιωμένες Εικόνες

Πολλές φορές οι πραγματικές εικόνες μικροσυστοιχιών δεν επαρκούν για να καλύψουν την μεγάλη ποικιλομορφία που μπορεί να συναντήσει κανείς στις εικόνες μικροσυστοιχιών. Για τον λόγο αυτό οι περισσότερες εργασίες αξιολογούνται κάνοντας χρήση και προσομοιωμένων εικόνων. Ο προσομοιωτής του Nykter [78] είναι ο περισσότερο διαδεδομένος προσομοιωτής, και ταυτόχρονα ο προσομοιωτής με τις

μεγαλύτερες δυνατότητες για την δημιουργία εικόνων μικροσυστοιχιών. Ο προσομοιωτής αυτός είναι αρθρωτός (modular), δηλαδή αποτελείται από ανεξάρτητες μονάδες που συνιστούν τμήματα μιας δομής ή συστήματος. Η κάθε μία από τις μονάδες είναι ανεξάρτητη και είτε μπορεί να μην χρησιμοποιηθεί κατά την προσομοίωση είτε μπορεί να αντικατασταθεί. Το σχήμα 2.13 παρουσιάζει το διάγραμμα των μονάδων του προσομοιωτή.



Σχήμα 2.14: Διάγραμμα με τις μονάδες του προσομοιωτή του Nykter.

Η μονάδα της εισαγωγής αρχείου είναι υπεύθυνη για την μετατροπή των εισαγόμενων δεδομένων στην διαμόρφωση που χρησιμοποιεί ο προσομοιωτής. Ένα αρχείο εισαγωγής συνήθως περιλαμβάνει τους λόγους του πράσινο με το κόκκινο κανάλι για κάθε έναν ανιχνευτή της εικόνας, ωστόσο μπορεί να περιέχει και τις θέσεις των ανιχνευτών στην μικροσυστοιχία, αλλά και διάφορες ακόμα πληροφορίες για τον κάθε ανιχνευτή. Αν τα παραπάνω δεν περιλαμβάνονται τότε ο προσομοιωτής δημιουργεί μόνος του αυτά τα δεδομένα.

Η πιο σημαντική μονάδα για την δημιουργία ρεαλιστικών δεδομένων μικροσυστοιχιών είναι η μονάδα που διαχειρίζεται τον βιολογικό θόρυβο των δεδομένων. Η μονάδα αυτή είναι υπεύθυνη για τον καθορισμό των στατιστικών χαρακτηριστικών των δεδομένων, που αφορούν κυτταρικές διεργασίες, αστοχίες κατά την προετοιμασία του πειράματος κ.α.

Η μονάδα κατασκευής της μικροσυστοιχίας είναι εκείνη, η οποία αποφασίζει για την χωροταξική δομή των κηλίδων επάνω στην εικόνα. Ο προσομοιωτής ενδέχεται να έχει ήδη την πληροφορία αυτή, αν του έχουν δοθεί οι θέσεις των κηλίδων από την μονάδα εισαγωγής αρχείου. Σε αυτήν την περίπτωση η συγκεκριμένη μονάδα παρακάμπτεται και ο

προσομοιωτής χωροθετεί τις κηλίδες ανάλογα με το αρχείο εισόδου. Ωστόσο, αν αυτή η πληροφορία δεν υπάρχει στο αρχείο εισόδου ο προσομοιωτής πρέπει να χωροθετήσει σε γραμμές και στήλες τις κηλίδες της εικόνας. Η συγκεκριμένη μονάδα είναι εκείνη η οποία προσθέτει σφάλματα στις θέσεις των κηλίδων αυξάνει η μειώνει την διακύμανση του μεγέθους των κηλίδων, και γενικά διαμορφώνει την χωροταξία της εικόνας με μια σειρά από παραμέτρους.

Διάφορα σφάλματα στην υβριδοποίηση των δειγμάτων με τους ανιχνευτές προσομοιώνονται από την μονάδα της υβριδοποίησης. Αυτά τα σφάλματα μπορεί να έχουν αντίκτυπο στο σχήμα των κηλίδων. Έτσι ο προσομοιωτής του Nykter δίνει την δυνατότητα να δημιουργήσει κηλίδες με κυκλικό σχήμα, αλλά και κηλίδες με ελλειπτικό σχήμα. Σε αυτήν την κατηγορία των σφαλμάτων συμπεριλαμβάνεται και η εκπομπή από την ίδια την μικροσυστοιχία, που έχει ως αποτέλεσμα την αύξηση του θορύβου υποβάθρου στην εικόνα. Τέλος ο προσομοιωτής προσθέτει τεχνουργήματα από γρατζουνιές στο υπόστρωμα ή από δείγμα που έχει χυθεί εκτός των θέσεων των ανιχνευτών.

Η μονάδα της σάρωσης προσομοιώνει αντίστοιχα την διαδικασία της σάρωσης της μικροσυστοιχίας ώστε να παραχθεί η εικόνα. Αυτή η διαδικασία μπορεί να εισάγει σφάλματα στην εικόνα, τα οποία έχει προβλέψει ο προσομοιωτής να φαίνονται στην εικόνα. Το πιο χαρακτηριστικό παράδειγμα αφορά το γεγονός ότι σάρωση ενδέχεται να γίνει υπό μικρή γωνία, με αποτέλεσμα η μικροσυστοιχία να φαίνεται λίγο στραμμένη ως προς τα όρια της ίδιας της εικόνας.

Ο προσομοιωτής τέλος έχει και μια μονάδα η οποία προσομοιώνει την διαδικασία της επεξεργασίας της εικόνας, η οποία ονομάζεται μονάδα ανάγνωσης της εικόνας. Όλες η παραπάνω μονάδες διαχειρίζονται από τον χρήστη με την βοήθεια ενός μεγάλου αριθμού παραμέτρων. Όλες αυτές οι παράμετροι είναι ομαδοποιημένες σε 5 ομάδες ρυθμίσεων ανάλογα με την μονάδα την οποία επηρεάζουν. Οι 5 ομάδες ρυθμίσεων είναι:

- Οι γενικές ρυθμίσεις του προσομοιωτή, από τις οποίες καθορίζονται τα γενικά χαρακτηριστικά των εικόνων και των δεδομένων που θα προσομοιωθούν. Για παράδειγμα από της γενικές ρυθμίσεις ορίζεται αν ο προσομοιωτής θα

προσομοιώσει μια cDNA μικροσυστοιχία ή μια μικροσυστοιχία της εταιρίας Affymetrix.

- Οι ρυθμίσεις θορύβου, οι οποίες καθορίζουν όλα τα είδη του θορύβου που μπορούν να επηρεάσουν την εξαγωγή της εικόνας μιας μικροσυστοιχίας.
- Οι ρυθμίσεις κατασκευής, οι οποίες καθορίζουν την τελική χωροταξική δομή της μικροσυστοιχίας.
- Οι ρυθμίσεις υβριδοποίησης, οι οποίες καθορίζουν την ποιότητα της μικροσυστοιχίας και κατ' επέκταση της εικόνας.
- Οι ρυθμίσεις σάρωσης, οι οποίες περιλαμβάνουν παραμέτρους που ρυθμίζουν τις φωτεινότητες της τελικής εξαγόμενης εικόνας, της υφής τους κτλ.

Λόγω του μεγάλου αριθμού των παραμέτρων, ο ίδιος ο προσομοιωτής προτείνει συγκεκριμένες τιμές για μια σειρά από παραμέτρους με σκοπό να δημιουργήσει τρεις κλιμακώσεις στην ποιότητας των εικόνων: καλές (good), κανονικές (normal), και κακές (bad).

2.4.Συμπεράσματα

Από τα εδάφια τα οποία προηγήθηκαν εύκολα καταλήγουμε στο συμπέρασμα ότι το πεδίο των μικροσυστοιχιών είναι πολύπλοκο. Όλα τα επιμέρους στάδια έχουν αναπτυχθεί πολύπλευρα, γεγονός το οποίο έχει άμεση επίπτωση στα επόμενα στάδια και συγκεκριμένα στα στάδια που απαιτούν ανάπτυξη λογισμικού. Έτσι και η επεξεργασία της εικόνας των μικροσυστοιχιών καλείται να καλύψει την ποικιλομορφία των μικροσυστοιχιών, που έχει προκύψει από της διαφορετικές κατασκευαστικές και βιολογικές προσεγγίσεις. Από τα εξειδικευμένα λογισμικά πακέτα του παρελθόντος πρέπει περάσουμε σε γενικευμένες μεθοδολογίες και σουίτες εργαλείων.

ΚΕΦΑΛΑΙΟ 3^ο: Βιβλιογραφική ανασκόπηση

3.Βιβλιογραφική ανασκόπηση

3.1.Εισαγωγή στις εικόνες μικροσυστοιχιών

Στο κεφάλαιο αυτό θα παρουσιαστεί η βιβλιογραφική ανασκόπηση στο πεδίο της επεξεργασία των εικόνων των μικροσυστοιχιών. Σκοπός είναι να περιγραφούν όλες οι εργασίες που έχουν υλοποιηθεί σε όλα τα στάδια της επεξεργασίας των εικόνων, από την εξαγωγή τους από το στάδιο της σάρωσης έως και την παραγωγή των δεδομένων.

Στην ιδανική περίπτωση, μια εικόνα μικροσυστοιχιών πρέπει να έχει τις ακόλουθες ιδιότητες [38,79]:

- όλες οι συστάδες να έχουν το ίδιο μέγεθος,
- τα διαστήματα μεταξύ των συστάδων να είναι ίσα,
- η απόσταση μεταξύ κηλίδων να είναι η ίδια,
- η θέση των κηλίδων πρέπει να είναι κεντροθετημένες σωστά στα σημεία που ορίζει το εκάστοτε πλέγμα,

- η μορφή των κηλίδων πρέπει να είναι τέλεια κυκλική και το ίδιο πράγμα πρέπει να συμβαίνει για όλες τις κηλίδες της εικόνας,
- δεν πρέπει να υπάρχουν μολύνσεις και άλλα αντικείμενα στην μικροσυστοιχία, τα οποία θα προκαλέσουν τεχνουργήματα στην παραγόμενη εικόνα
- πρέπει να υπάρχει ελάχιστη και ομοιόμορφη ένταση του θορύβου υποβάθρου στην εικόνα.

Εάν όλοι αυτοί οι εξιδανικευμένοι όροι ικανοποιούνταν, ο στόχος της επεξεργασίας εικόνας θα μπορούσε να ολοκληρωθεί εύκολα από ένα απλό λογισμικό [80]. Ένα πρόγραμμα, το οποίο θα προσομοίωνε μια σειρά κυκλικών δίσκων με καθορισμένες διαστάσεις και καθορισμένα διαστήματα μεταξύ τους, θα μπορούσε να υπερτεθεί της εικόνα εντοπίζοντας όλες της κηλίδες της. Τα εικονοστοιχεία που θα βρίσκονταν μέσα σε αυτούς τους κυκλικούς δίσκους θα θεωρούνταν εικονοστοιχεία σήματος ενώ εκείνα που θα βρίσκονταν στο εξωτερικό των κυκλικών δίσκων θα ορίζονταν ως εικονοστοιχεία υποβάθρου. Εντούτοις, σχεδόν όλοι οι παραπάνω όροι παραβιάζονται στις πραγματικές εικόνες μικροσυστοιχιών. Οι πραγματικές εικόνες παρουσιάζουν μια σειρά από προβλήματα, τα οποία καλούνται να παρακάμψουν οι μέθοδοι επεξεργασίας εικόνας. Τέτοια προβλήματα είναι η μικρή μετατόπιση των κηλίδων από τις κεντροθετημένες θέσεις τους, η απόκλιση της μορφής των κηλίδων από το κυκλικό σχήμα, τα διαφορετικά μεγέθη των κηλίδων μέσα στην ίδια εικόνα, και τέλος ο θόρυβος στις εικόνες, ο οποίος παρουσιάζεται είτε υπό την μορφή τεχνουργημάτων, είτε υπό την μορφή θορύβου υποβάθρου.

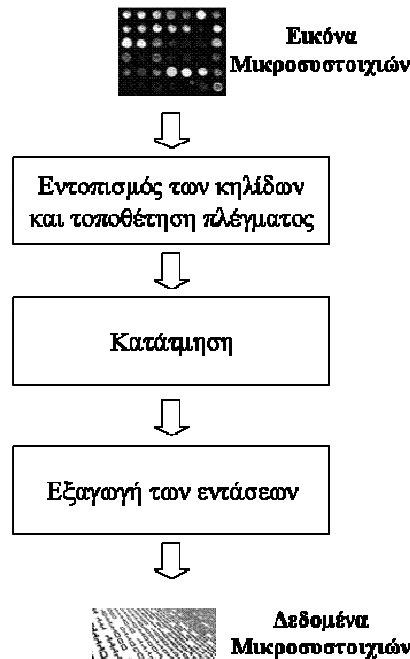
Αναλύοντας τους λόγους για τους οποίους δεν ισχύουν οι παραπάνω όροι στις πραγματικές εικόνες, διαπιστώνεται ότι οι περισσότεροι από αυτούς παραβιάζονται λόγω των μηχανικών περιορισμών κατά την διαδικασία της εκτύπωσης των μικροσυστοιχιών. Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο, όσο αναπτύσσεται η τεχνολογία των μικροσυστοιχιών αυξάνεται και η απαίτηση για περισσότερους ανιχνευτές επάνω σε μια μικροσυστοιχία. Το γεγονός αυτό όμως είναι αντικρουόμενο με το θεμελιώδες πρόβλημα στην κατασκευή μίας μικροσυστοιχίας το οποίο είναι η διασφάλιση ότι όλες οι κηλίδες είναι χωρικά διακριτές μέσα στην παραγόμενη εικόνα. Οι αδυναμία των ρομποτικών συστημάτων να κρατήσουν την ισορροπία στις δύο αυτές απαιτήσεις, έχουν επιπτώσεις

στην ποιότητα της κατασκευής. Συγκριμένα, υπάρχουν διάφοροι παράγοντες που συντελούν στην απόκλιση του σχήματος των κηλίδων από το κυκλικό, αλλά και στην διαφορετικότητα των μεγεθών των κηλίδων σε μια μικροσυστοιχία. Μερικοί από αυτούς είναι:

- το γεγονός ότι τα μεγέθη των σταγονιδίων ενός διαλύματος DNA (κυρίως όταν χρησιμοποιούνται ακίδες κατά την εκτύπωση) μπορούν να ποικίλουν
- οι συγκεντρώσεις του DNA και του άλατος στο διάλυμα μπορούν να αλλάξουν με τον χρόνο, κάνοντας τη μορφή των κηλίδων να αποκλίνει από το κυκλικό
- η πυκνότητα του διαλύματος που τοποθετείται σε έναν ανιχνευτή μπορεί να είναι ανομοιόμορφη

Ένα άλλο πρόβλημα των πραγματικών εικόνων, για το οποίο πρέπει να γίνει ιδιαίτερη αναφορά, είναι ο θόρυβος που παρουσιάζουν οι εικόνες και οι αιτίες που τον προκαλούν. Ο θόρυβος στις εικόνες εμφανίζεται είτε υπό την μορφή τεχνουργημάτων είτε υπό την μορφή θορύβου υποβάθρου. Τα τεχνουργήματα δημιουργούνται από την αερομεταφερόμενη σκόνη, η οποία μπορεί να μολύνει τη μικροσυστοιχία δημιουργώντας σήμα φθορισμού στις σαρωμένες εικόνες. Επίσης τεχνουργήματα μπορούν να δημιουργηθούν από ακαθαρσίες και γρατσουνιές στην επιφάνεια του υποστρώματος προκαλώντας στίγματα στην εικόνα. Φωτεινά αντικείμενα μπορούν τέλος να δημιουργηθούν από την υπερχειλίση των δειγμάτων έξω από τις θέσεις των ανιχνευτών. Από την άλλη μεριά, ο θόρυβος υποβάθρου δημιουργείται είτε από την εκπομπή του ίδιου το υποστρώματος είτε από ανακλάσεις της ηλεκτρομαγνητικής ακτινοβολίας κατά την σάρωση της μικροσυστοιχίας.

Όλα τα παραπάνω προβλήματα των εικόνων μπορούν να προκαλέσουν σφάλματα στα αποτελέσματα και για τον λόγο αυτό πρέπει να αντιμετωπιστούν από τις μεθόδους επεξεργασίας των εικόνων ώστε να μην έχουν επιπτώσεις στην ποιότητα των εξαγόμενων δεδομένων. Για την αντιμετώπιση όλων αυτών των φαινομένων οι διαδικασίες της επεξεργασίας εικόνας αναπτύσσεται σε τρία στάδια [81-83], όπως φαίνεται στο παρακάτω διάγραμμα του σχήματος 3.1.



Σχήμα 3.1: Γενικό διάγραμμα ροής της επεξεργασίας των εικόνων των μικροσυστοιχιών

Όπως φαίνεται στο σχήμα 3.1 τα στάδια της επεξεργασίας των εικόνων των μικροσυστοιχιών είναι:

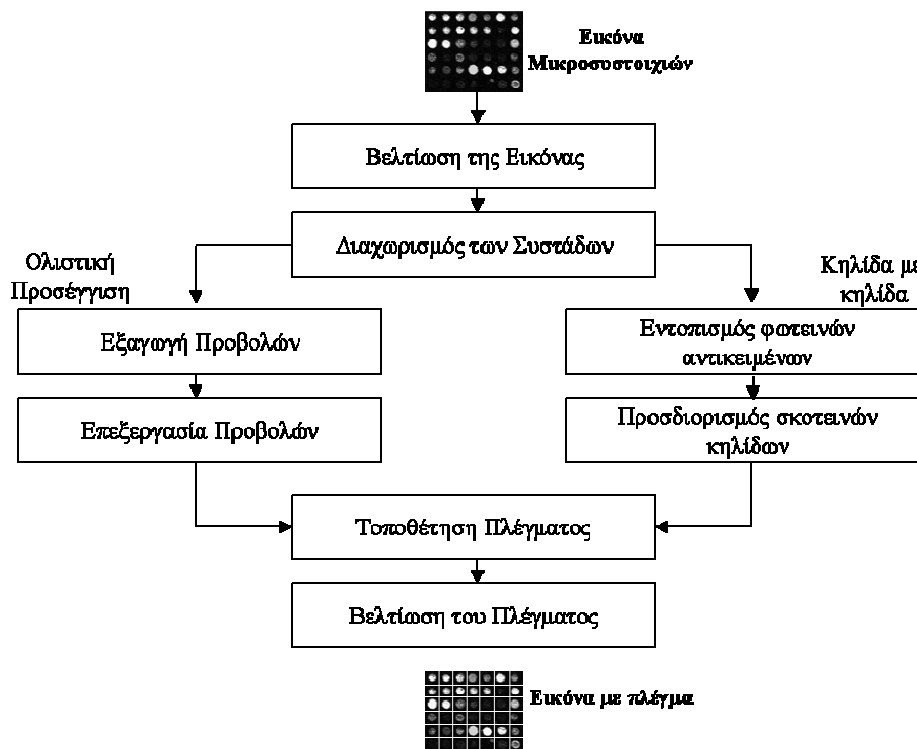
- Ο εντοπισμός των κηλίδων και η τοποθέτηση πλέγματος (spot addressing and gridding) που στοχεύει στον προσδιορισμό της θέσης κάθε κηλίδας στην εικόνα, και στην υπέρθεση ενός πλέγματος επάνω σε αυτή με σκοπό κάθε κηλίδα να απομονωθεί σε μια μικρή περιοχή της εικόνας.
- Η κατάτμηση (segmentation), η οποία συντελείται ξεχωριστά για κάθε κηλίδα και έχει σκοπό τον διαχωρισμό των εικονοστοιχείων σήματος από τα εικονοστοιχεία υποβάθρου.
- Η εξαγωγή της έντασης (intensity extraction), η οποία στοχεύει στην ποσοτικοποίηση της υβριδοποίησης μεταξύ των ανιχνευτών και των δειγμάτων.

3.2.Εντοπισμός των κηλίδων και τοποθέτηση πλέγματος

Ο θεμελιώδης στόχος της επεξεργασίας μίας εικόνας μικροσυστοιχιών είναι η μέτρηση της έντασης των παρατεταγμένων κηλίδων ώστε να ποσοτικοποιηθούν τα επίπεδα της υβριδοποίησης που αντιστοιχούν σε αυτά. Πιο συγκεκριμένα, η επεξεργασία της εικόνας

αλλά και των δεδομένων των μικροσυστοιχιών, θα αξιολογήσει και την αξιοπιστία της ποσοτικοποίησης των κηλίδων και θα παρέχει τις προειδοποιήσεις για πιθανά προβλήματα κατά τη διάρκεια της διαδικασίας κατασκευής και υβριδοποίησης της μικροσυστοιχίας. Προτού να εκτελεστούν όμως αυτές οι διαδικασίες, προηγείται το βήμα του εντοπισμού των θέσεων των κηλίδων μέσα στην εικόνα. Παρακάτω θα ακολουθήσει μια συνοπτική περιγραφή των υπαρχουσών προσεγγίσεων που χρησιμοποιούνται μέχρι σήμερα για τον εντοπισμό των κηλίδων.

Ο εντοπισμός των κηλίδων είναι εν γένει μία δύσκολη διαδικασία, της οποίας τα βήματα είναι συγκεχυμένα. Όλοι δηλαδή οι ερευνητές δεν ακολουθούν μια τυποποιημένη διαδικασία. Ωστόσο, αν κανείς ανατρέξει σε ολόκληρη την βιβλιογραφία θα διαπιστώσει ότι υπάρχουν συγκεκριμένα βήματα, τα οποία όμως οι ερευνητές τα ακολουθούν κατά περίπτωση. Τα βήματα δε αυτά αν συγκεντρωθούν, μπορεί να εξαχθεί μια χρονική σειρά με την οποία εφαρμόζονται. Τα διάγραμμα ροής του σχήματος 3.2 παρουσιάζει συγκεντρωμένα όλα τα βήματα τα οποία οι ερευνητές κατά περίπτωση χρησιμοποιούν για να ολοκληρώσουν την διαδικασία του εντοπισμού των κηλίδων στην εικόνα.



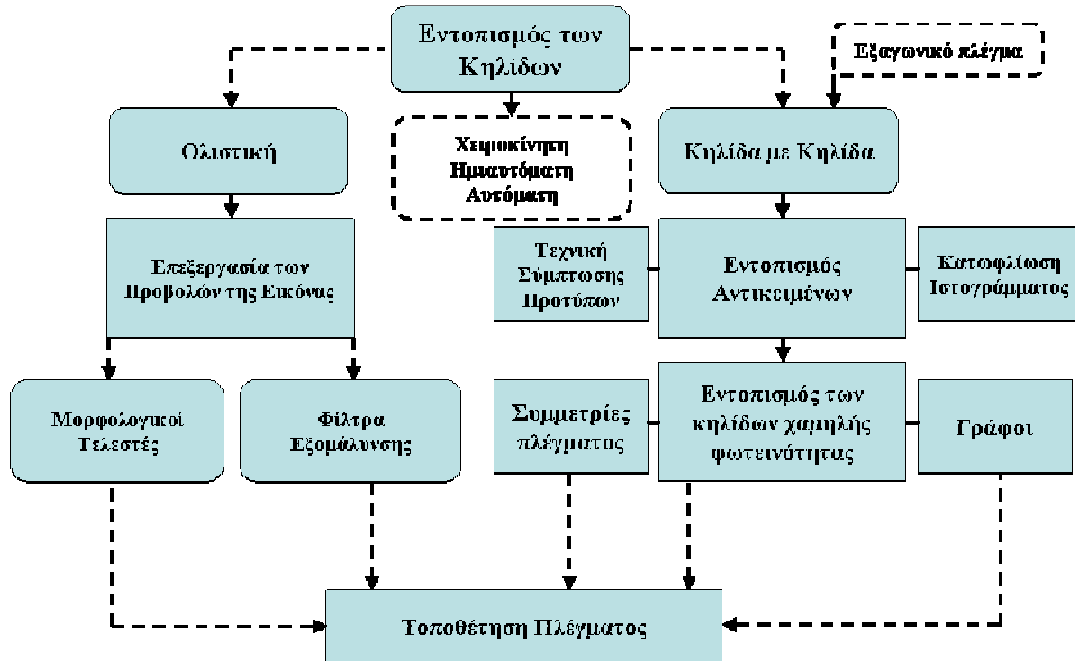
Σχήμα 3.2: Διάγραμμα ροής του εντοπισμού των κηλίδων

Όπως φαίνεται στο σχήμα 3.2 πολλές από τις εργασίες ξεκινούν με την βελτίωση των εικόνων με σκοπό να απαλείψουν τον όποιο θόρυβο στην εικόνα. Με αυτό τον τρόπο θα γίνει πολύ πιο εύκολη η περαιτέρω επεξεργασία για τον εντοπισμό των κηλίδων. Οι περισσότερες εργασίες όμως δεν εφαρμόζουν το βήμα αυτό διότι από μόνη της η βελτίωση μεγάλων εικόνων είναι μια χρονοβόρα διαδικασία και απαιτεί πολλούς πόρους του συστήματος. Έτσι το ισοζύγιο μεταξύ των πλεονεκτημάτων μιας βελτιωμένης εικόνας και των υψηλών απαιτήσεων από το σύστημα, ωθεί τους ερευνητές να παρακάμπτουν το βήμα αυτό. Αντίθετα το βήμα του διαχωρισμού των συστάδων είναι απαραίτητο. Κάθε συστάδα από κηλίδες πρέπει να επεξεργαστεί ξεχωριστά, και άρα μια μέθοδος εντοπισμού των κηλίδων είτε χρειάζεται έναν αυτόματο διαχωρισμό είτε θα δέχεται ως είσοδο τις επιμέρους συστάδες των εικόνων. Τα επόμενα βήματα διαχωρίζονται ανάλογα με την προσέγγιση που επιλέγεται. Όπως θα δούμε και παρακάτω οι μέθοδοι τις βιβλιογραφίας χωρίζονται σε δύο μεγάλες κατηγορίες: (α) την ολιστική προσέγγιση, η οποία δημιουργεί το πλέγμα απευθείας πάνω στην εικόνα χωρίς να λαμβάνει υπόψη τις θέσεις της κάθε κηλίδας στην εικόνα, και (β) την προσέγγιση κηλίδα με κηλίδα όπου η κάθε κηλίδα αφού εντοπιστεί η θέση της στην εικόνα αντιστοιχίζεται με μια περιοχή, η οποία απομονώνει την κηλίδα από τις υπόλοιπες κηλίδες. Ως προ-τελευταίο βήμα σημειώνεται η τοποθέτηση του πλέγματος. Οι μέθοδοι που ακολουθούν την ολιστική προσέγγιση έχουν ως αυτοσκοπό το βήμα αυτό. Σκοπός τους δηλαδή είναι καθαυτή η δημιουργία ενός πλέγματος που να απομονώνει τις κηλίδες της εικόνας όσο το δυνατόν καλύτερα.. Αντίθετα η μέθοδοι που εντοπίζουν τις θέσεις της κάθε κηλίδας ξεχωριστά (προσέγγιση κηλίδα με κηλίδα) ενδέχεται να παραλείπουν το βήμα αυτό. Στην ουσία, και σε αυτήν την περίπτωση το βήμα αυτό δεν παραλείπεται αλλά απλώς δεν εφαρμόζεται κάποιος έξυπνος αλγόριθμος για την υλοποίησή του. Η εύκολη λύση σε αυτήν την περίπτωση είναι η τοποθέτηση ενός τετράγωνου κελιού γύρω από το εντοπισμένο κέντρο της κηλίδας. Τέλος πολλές από τις μεθόδους εισάγουν ένα τελευταίο βήμα κατά το οποίο βελτιώνεται το παραγόμενο πλέγμα. Αυτό γίνεται κυρίως από τις μεθόδους που ακολουθούν την ολιστική προσέγγιση, διότι στην ολιστική προσέγγιση οι επιμέρους θέσεις των κηλίδων δεν έχουν συμπεριληφθεί κατά την επεξεργασία και έτσι το πλέγμα χρειάζεται βελτίωση σε τοπικό επίπεδο για να αντιμετωπίσει τις αποκλίσεις των θέσεων των κηλίδων από τις κεντροθετημένες θέσεις τους.

Τα πρώτα χρόνια που άρχισαν να παρουσιάζονται εργασίες για την επεξεργασία των εικόνων των μικροσυστοιχιών η κατηγοριοποίηση των μεθόδων για το στάδιο του εντοπισμού των κηλίδων γίνονταν ανάλογα με την παρέμβαση του χρήστη στην διαδικασία [84]. Έτσι οι μέθοδοι χωρίζονται σε χειροκίνητες (manual), ημι-αυτόματες (semi-automated), και αυτόματες (automated). Οι χειροκίνητες μέθοδοι είχαν δύο βασικές απαιτήσεις από τους χρήστες. Αφενός απαιτούσαν μια σειρά από παραμέτρους των εικόνων και αφετέρου ο χρήστης ήταν αυτός που έπρεπε χειροκίνητα να μετακινήσει το παραγόμενο πλέγμα στην εικόνα έως ότου τοποθετηθεί στην κατάλληλη θέση. Μερικές από τις παραμέτρους που απαιτούνται είναι η διάμετρος των κηλίδων, ο αριθμός των συστάδων, ο αριθμός των κηλίδων σε κάθε συστάδα, ο αριθμός των κηλίδων σε κάθε γραμμή και στήλη από κηλίδες σε κάθε συστάδα, οι απόσταση μεταξύ δύο γειτονικών συστάδων, η απόσταση μεταξύ δύο γειτονικών συστάδων κ.α. Το πιο χαρακτηριστικό παράδειγμα αυτής της κατηγορίας είναι το πρόγραμμα ScanAlyze [75]. Οι ημι-αυτόματες μέθοδοι [85] αναφέρονται στις μεθόδους, οι οποίες απαιτούν από τον χρήστη τις τιμές μόνο των παραμέτρων, πραγματοποιώντας όλες τις υπόλοιπες διαδικασίες αυτόματα. Τέλος οι αυτόματες μέθοδοι δεν απαιτούν καμία πληροφορία ή παρέμβαση από τον χρήστη.

Κατηγοριοποιώντας όλες τις μεθόδους που έχουν παρουσιαστεί στην βιβλιογραφία για το στάδιο του εντοπισμού των κηλίδων, προκύπτει το δένδρογραμμα του σχήματος 3.3. Όπως αναφέρθηκε και ανωτέρω οι μέθοδοι κατηγοριοποιούνται σε δύο κατηγορίες, την ολιστική προσέγγιση και την προσέγγιση κηλίδα με κηλίδα. Η ολιστική προσέγγιση εν γένει χρησιμοποιεί τις προβολές της εικόνας με σκοπό να εξάγει τις παραμέτρους που απαιτούνται για την δημιουργία του πλέγματος. Ωστόσο υπάρχουν και ολιστικές μέθοδοι, οι οποίες δημιουργούν το πλέγμα χωρίς την αξιοποίηση των προβολών. Από την άλλη μεριά η μέθοδοι «κηλίδα με κηλίδα» αντιμετωπίζουν δύο βασικά βήματα, εφόσον σκοπός τους είναι να προσδιορίσουν τις θέσεις όλων των κηλίδων (είτε υψηλής είτε χαμηλής φωτεινότητας). Έτσι πρώτα εντοπίζουν τα φωτεινά αντικείμενα της εικόνας με απλές μεθόδους κατάτμησης, και εν συνεχεία προσδιορίζουν τις θέσεις όπου πιθανόν υπάρχουν κηλίδες χαμηλής φωτεινότητας που δεν εντοπίζονται από το προηγούμενο βήμα. Στο σχήμα επισημαίνεται επίσης η αρχική κατηγοριοποίηση των μεθόδων σε χειροκίνητες, ημι-αυτόματες και αυτόματες, αλλά και το γεγονός ότι σε περίπτωση που οι κηλίδες στην

εικόνα είναι χωροθετημένες με βάση εξαγωνικό πλέγμα, οι μέθοδοι για τον αυτόματο εντοπισμό των κηλίδων ακολουθεί την προσέγγιση κηλίδα με κηλίδα.



Σχήμα 3.3: Δενδρόγραμμα των μεθόδων για τον εντοπισμό των κηλίδων

Παρακάτω θα ακολουθηθεί το διάγραμμα ροής του σχήματος 3.2 με σκοπό να αναπτυχθούν ένα-ένα όλα τα επιμέρους βήματα που έχουν χρησιμοποιηθεί από τις εργασίες της βιβλιογραφίας για το στάδιο του εντοπισμού των κηλίδων.

3.2.1.Βελτίωση της εικόνας

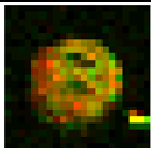
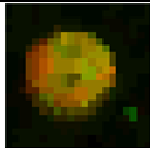
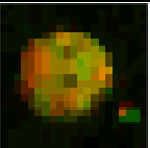
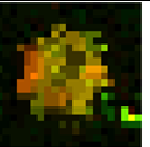
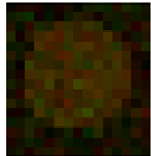
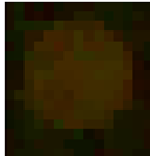
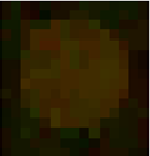
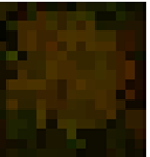
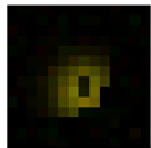

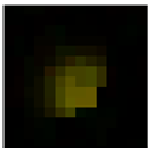

Σκοπός της βελτίωσης της εικόνας είναι η απαλοιφή κάθε είδους θορύβου από την εικόνα. Ο θόρυβος στην εικόνα των μικροσυστοιχιών παρουσιάζεται με την μορφή είτε τεχνουργημάτων είτε θορύβου υποβάθρου. Ιδιαίτερα τα τεχνουργήματα στην εικόνα μπορούν να προκαλέσουν προβλήματα στον εντοπισμό των κηλίδων διότι πολλές φορές αναγνωρίζονται από τις μεθόδους ως κηλίδες. Αναμφισβήτητα λοιπόν η επεξεργασία μιας εικόνας απαλλαγμένης από τον θόρυβο των τεχνουργημάτων είναι ευκολότερη από τον εντοπισμό των κηλίδων στην αρχική γεμάτη θόρυβο εικόνα. Ωστόσο οι μέθοδοι που

αναπτύσσονται για τον εντοπισμό των κηλίδων μπορούν από μόνες τους να παρακάμπτουν σε μεγάλο βαθμό την ύπαρξη των τεχνουργημάτων δεδομένου ότι εκμεταλλεύονται τις συμμετρίες της εικόνας, γνωρίζοντας σε ποιες θέσεις κατά προσέγγιση αναμένεται να υπάρχει κηλίδα. Το γεγονός αυτό σε συνδυασμό με τις απαιτήσεις σε χρόνο και σε πόρους του συστήματος για την βελτίωση της εικόνας, οδηγεί τους ερευνητές στην αποφυγή του βήματος αυτού στην ροή των μεθόδων τους.

Διάφορα φίλτρα έχουν χρησιμοποιηθεί για την βελτίωση της εικόνας. Συγκεκριμένα στο πεδίο των μικροσυστοιχιών έχει γίνει χρήση του φίλτρου διαμέσου (median filter) [86], του μεταβλητού φίλτρου διαμέσου (adaptive median filter) [87], αλλά και του red hat φίλτρου [86-87]. Επίσης για την βελτίωση της εικόνας έχουν χρησιμοποιηθεί μορφολογικοί τελεστές (morphological operators) [88] αλλά και διάφοροι μετασχηματισμοί όπως ο Hough μετασχηματισμός [89] και ο στατικό κυματοειδής μετασχηματισμός (Stationary Wavelet Transform) [54]. Όλα τα παραπάνω φίλτρα και οι μορφολογικοί τελεστές εφαρμόζονται κατά την διάρκεια της προεπεξεργασίας της εικόνας και στα δύο κανάλια της. Την καινοτομία στην βελτίωση της εικόνας έκαναν το 2004 οι Lukac *et al.* [90-91] οι οποίοι πρότειναν μια σειρά από φίλτρα τα οποία βελτιώνουν πολυκαναλικές εικόνες. Τα φίλτρα αυτά δηλαδή δεν εφαρμόζονται σε κάθε ένα κανάλι ξεχωριστά αλλά απευθείας στην πολυκαναλική εικόνα. Δεδομένου ότι και εικόνες των μικροσυστοιχιών είναι πολυκαναλικές εικόνες (έχουν δύο κανάλια) οι Lukac *et al.* για την αξιολόγηση των φίλτρων τους χρησιμοποίησαν και εικόνες μικροσυστοιχιών με τα αποτελέσματα μάλιστα να παρουσιάζουν ιδιαίτερο ενδιαφέρον. Τα φίλτρα αυτά λέγονται: φίλτρα διάμεσου διανύσματος (Vector Median Filter – VMF), σταθμισμένα φίλτρα διάμεσου διανύσματος (Weighted Vector Median Filters – WVMF), και φίλτρο διεύθυνσης βασικού διανύσματος (Basic Vector Directional Filter). Στο σχήμα 3.4 παρουσιάζονται τα αποτελέσματα των φίλτρων αυτών για τρεις διαφορετικές κηλίδες. Όπως φαίνεται και από το σχήμα, τα φίλτρα δρουν αποτελεσματικά στην εικόνα των μικροσυστοιχιών εξομαλύνοντας τα τεχνουργήματα και τον κρουστικό θόρυβο (impulse noise).

Εκτός από την εξομάλυνση της εικόνας, το βήμα της βελτίωσης των εικόνων των μικροσυστοιχιών έχει έναν ακόμα ρόλο. Αυτός είναι η περιστροφή της εικόνας.

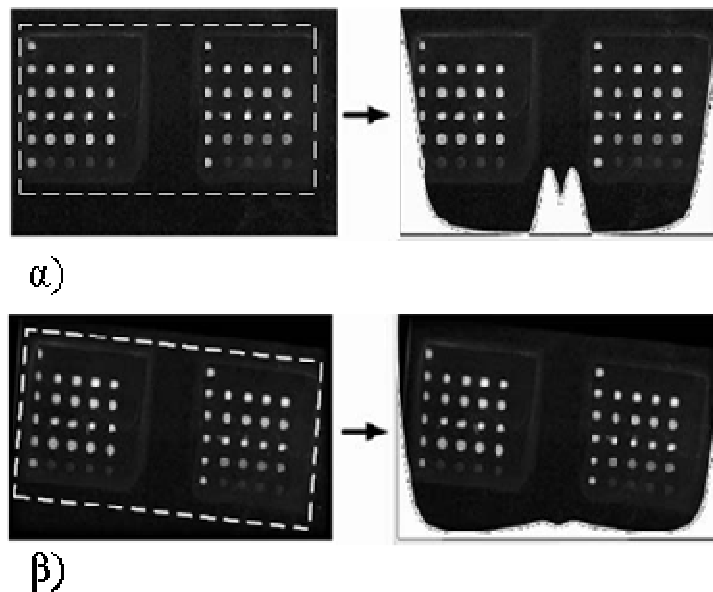
Ειδικότερα τα πρώτα χρόνια, όπου το υλικό του πειράματος δεν είχε εξελιχτεί, στις εικόνες παρατηρούταν συχνά το φαινόμενο η μικροσυστοιχία να μην είναι αποτυπωμένη παράλληλα με την εικόνα, αλλά να παρουσιάζει μικρή στρέψη. Το γεγονός αυτό δυσκόλευε τις αυτόματες μεθόδους επεξεργασίας, η οποίες υλοποιούνταν με σκοπό να επεξεργαστούν μια συμμετρική εικόνα όπου οι κηλίδες είναι τοποθετημένες σε κατακόρυφες στήλες και οριζόντιες γραμμές. Για τον λόγο αυτό, πολλές μέθοδοι σε αυτό το αρχικό βήμα διορθώνουν την γωνία στρέψης της μικροσυστοιχίας ως προς τα όρια της εικόνας.

| | Αρχική | VMF | WVMF | VBDF |
|------------------------------------|---|---|--|---|
| Κηλίδα με τεχνούργημα |  |  |  |  |
| Κηλίδα χαμηλής φωτεινότητας |  |  |  |  |
| Κηλίδα με σχήμα (doughnut) |  |  |  |  |

Σχήμα 3.4: Το αποτέλεσμα των πολυκαναλικών φίλτρων σε τρεις διαφορετικές κηλίδες

Για να το επιτύχουν αυτό οι περισσότερες μέθοδοι εντοπίζουν της τέσσερις γωνίες των συστάδων [92], υπολογίζουν την κλίση μεταξύ τους και στρέφουν αντίστοιχα την εικόνα. Η συγκεκριμένη τεχνική βέβαια αντιμετωπίζει προβλήματα όταν στις τέσσερις άκρες της συστάδας υπάρχουν σκοτεινές κηλίδες, οι οποίες δεν μπορούν να εντοπιστούν εύκολα. Έτσι οι Deng και Duan [93] πρότειναν μία μέθοδο, η οποία είναι βασισμένη στο φάσμα ισχύος (Power Spectra) για να υπολογίσουν την στρέψη που έχει η μικροσυστοιχία σε σχέση με την εικόνα. Στο σχήμα 3.5 παρουσιάζεται η τρόπος με τον οποίο η μέθοδος αυτή εντοπίζει, αν υπάρχει, την κλίση μεταξύ μικροσυστοιχίας και εικόνας.

Η μέθοδος εξάγει το φάσμα ισχύος της εικόνας και από αυτό προσδιορίζει την γωνία κατά την οποία είναι στραμμένη η μικροσυστοιχία σε σχέση με τα όρια της εικόνας. Τα τελευταία χρόνια, όπου ο εξοπλισμός του πειράματος των μικροσυστοιχιών έχει εξελιχθεί σημαντικά, οι εικόνες σπάνια έχουν πρόβλημα στρέψης και έτσι οι νέες μέθοδοι δεν συμπεριλαμβάνουν κανένα αλγόριθμο διόρθωσης της κλίσης της μικροσυστοιχίας.

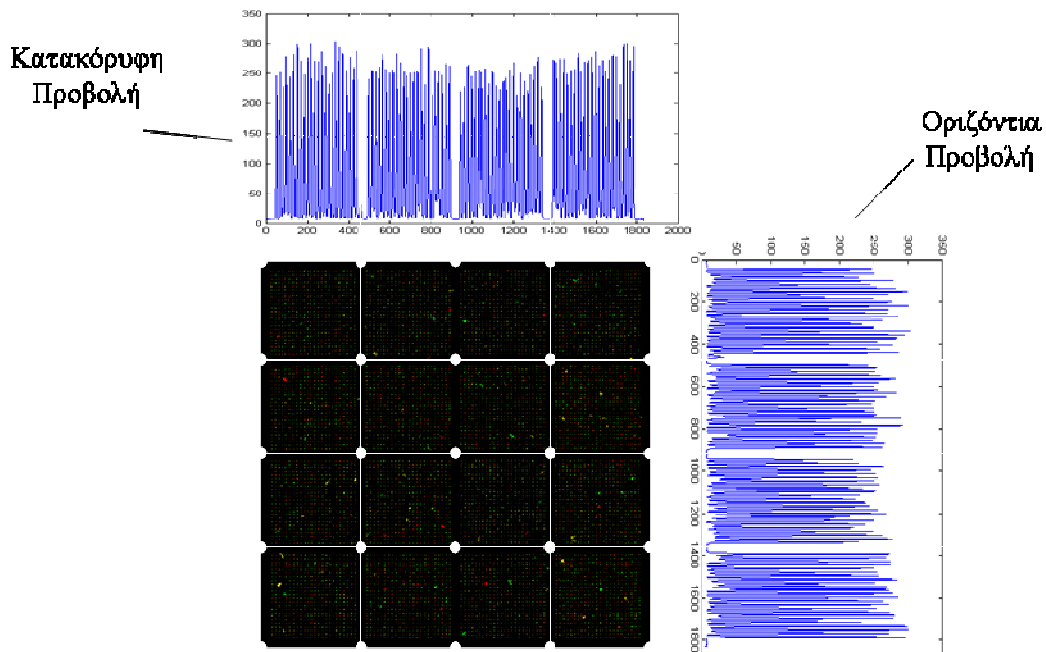


Σχήμα 3.5: Διόρθωση της στρέψης των εικόνων μικροσυστοιχιών υπολογίζοντας το φάσμα ισχύος τις εικόνας.: α) εικόνα χωρίς στρέψη, β) στραμμένη εικόνα.

3.2.2. Διαχωρισμός των συστάδων

Ο διαχωρισμός των συστάδων είναι ένα βήμα, το οποίο θα έπρεπε να συμπεριλαμβάνουν όλες οι αυτόματες μέθοδοι για τον εντοπισμό των κηλίδων. Είναι αναγκαίο η επεξεργασία των εικόνων να γίνεται για κάθε μία συστάδα ξεχωριστά, γεγονός το οποίο σημαίνει ότι αν μια μέθοδος δεν έχει συμπεριλάβει το βήμα διαχωρισμού των συστάδων ή θα δέχεται ως είσοδο μικροσυστοιχίες με μόνο μια συστάδα, ή θα πρέπει ο διαχωρισμός της εικόνας με περισσότερες συστάδες να γίνεται χειροκίνητα από τον χρήστη. Είναι ωστόσο γεγονός ότι το βήμα αυτό είναι ίσως το πιο εύκολο βήμα της επεξεργασία των εικόνων, και για τον λόγο αυτό δεν είναι στο επίκεντρο της έρευνας. Ο διαχωρισμός των συστάδων γίνεται με την χρήση των προβολών της εικόνας [94-96]. Οι προβολές της εικόνας είναι τα δύο

μονοδιάστατα σήματα, τα οποία προκύπτουν από το άθροισμα των φωτεινοτήτων των εικονοστοιχείων στην κατακόρυφη και την οριζόντια διεύθυνση. Συγκεκριμένα για την εξαγωγή της κατακόρυφης προβολής αθροίζονται οι εντάσεις των εικονοστοιχείων για κάθε στήλη εικονοστοιχείων της εικόνας. Αντίστοιχα η οριζόντια προβολή εξάγεται από το άθροισμα της κάθε γραμμής από εικονοστοιχεία στην εικόνα. Για τον διαχωρισμό των συστάδων εντοπίζονται οι μεγάλες κοιλάδες των δύο προβολών. Τα σημεία στα οποία βρίσκονται οι κοιλάδες αυτές υποδεικνύουν τις θέσεις που θα χωριστούν οι συστάδες. Από την κατακόρυφη προβολή εξάγονται οι τεταγμένες, ενώ από την οριζόντια προβολή οι τεταγμένες, των σημείων στα οποία θα χωριστεί εικόνα. Την διαδικασία αυτή παρουσιάζει η το σχήμα 3.6.

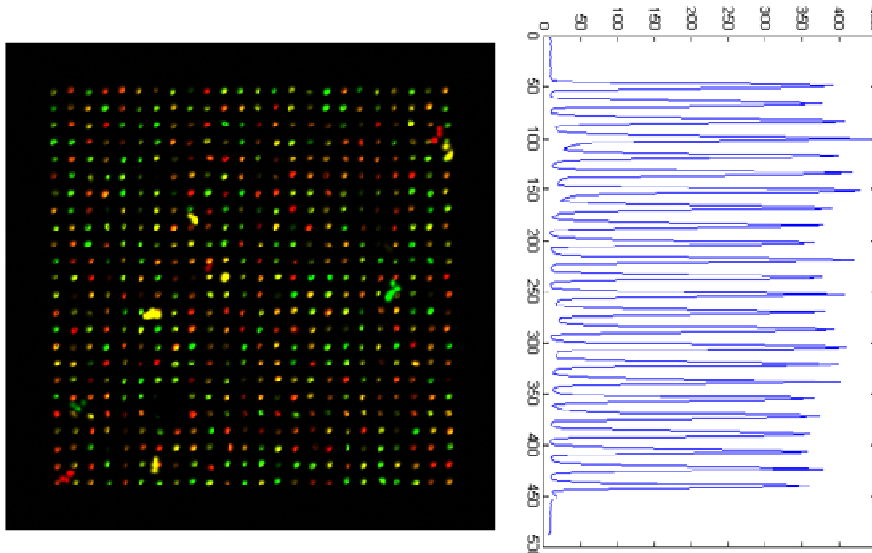


Σχήμα 3.6: Η κατακόρυφη και η οριζόντια προβολή μιας εικόνας με 16 συστάδες. Η ευθείες που ορίζονται από τις κοιλάδες των δύο προβολών χωρίζουν την εικόνα στις επιμέρους συστάδες, στα σημεία στα οποία τέμνονται.

Όπως θα δούμε στο επόμενο εδάφιο οι προβολές της εικόνας χρησιμοποιούνται κατά βάση στην ολιστική προσέγγιση, με σκοπό την εξαγωγή του πλέγματος

3.2.3.Ολιστική Προσέγγιση

Η ολιστική προσέγγιση αναφέρεται στις μεθόδους εκείνες, οι οποίες δημιουργούν ένα πλέγμα στα μέτρα της κάθε εικόνας, με σκοπό να απομονώσουν σε κάθε κελί του μια και μόνο κηλίδα. Για να το επιτύχουν αυτό οι αυτόματες μέθοδοι απαιτείται η εξεύρεση μιας σειράς από παραμέτρους, όπως το μέγεθος των κηλίδων, η απόσταση μεταξύ γειτονικών κηλίδων, ο αριθμός των κηλίδων κ.ο.κ.. Η πλειονότητα των μεθόδων [94-98] που ακολουθούν αυτήν την προσέγγιση χρησιμοποιούν τις προβολές της εικόνας. Η προβολές της εικόνας πέραν των μεγάλων κοιλάδων, οι οποίες δημιουργούνται από τα κενά μεταξύ γειτονικών συστάδων, παρουσιάζουν μικρότερες κοιλάδες οι οποίες δημιουργούνται ανάμεσα στις γραμμές και τις στήλες από κηλίδες, όπως δείχνει το σχήμα 3.7.



Σχήμα 3.7. Η οριζόντια προβολή μίας συστάδας.

Για τις περιπτώσεις όπου εικόνες χαρακτηρίζονται από πολύ θόρυβο οι προβολές επεξεργάζονται, όπως ακριβώς θα επεξεργαζόταν ένα μονοδιάστατο σήμα. Σκοπός της επεξεργασίας των προβολών είναι η έμφαση των κορυφών και των κοιλάδων του σήματος. Καθιστώντας τις κορυφές και τις κοιλάδες του σήματος περισσότερο εμφανείς οι μέθοδοι μπορούν τις εντοπίζουν αποτελεσματικότερα και με μεγαλύτερη αξιοπιστία. Η βελτίωση των προβολών γίνεται με χρήση διάφορων τεχνικών επεξεργασίας σήματος. Στην βιβλιογραφία έχουν εφαρμοστεί διάφοροι συνδυασμοί μορφολογικών πράξεων (morphological operators) [99-102], αλλά και εξομάλυνση των προβολών με χρήση

φίλτρων εξομάλυνσης (smoothing filters) [103-104]. Η μέθοδοι αυτές συνεχίζουν με την εξεύρεση του βέλτιστου πλέγματος, προσδιορίζοντας τις παραμέτρους που χρειάζονται.

Στην ολιστική προσέγγιση κατατάσσονται επίσης, ορισμένες εργασίες, οι οποίες δημιουργούν το πλέγμα χωρίς να εντοπίζουν μεμονωμένα τις θέσεις της κάθε κηλίδας, αλλά οι οποίες δεν κάνουν χρήση των προβολών της εικόνας για τον σκοπό αυτό. Η κυριότερη μεθοδολογία είναι αυτή των Bariamis *et. al.* [105-107], η οποία είναι βασισμένη στις Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines - SVMs). Πιο συγκεκριμένα, η μέθοδος αυτή αφού εντοπίσει τα φωτεινά αντικείμενα δημιουργεί τις βέλτιστες ευθείες μεταξύ γραμμών και στηλών των αντικειμένων αυτών εκμεταλλεύομενη την διαδικασία εκπαίδευσης των Μηχανών Διανυσμάτων Υποστήριξης, οι οποίες εξ ορισμού εντοπίζουν το βέλτιστο υπερεπίπεδο που διαχωρίζει ένα σύνολο δεδομένων σύμφωνα με συγκεκριμένα χαρακτηριστικά. Τέλος στην ολιστική προσέγγιση μπορεί να καταταγεί η εργασία των Zacharia *et. al.* [108-109] που διαχειρίζεται την τοποθέτηση πλέγματος ως ένα πρόβλημα βελτιστοποίησης το οποίο αντιμετωπίζει με χρήση γενετικών αλγόριθμων. Η μέθοδος αυτή χωρίζει την εικόνα σε λωρίδες (είτε οριζόντιες για τον εντοπισμό των γραμμών, είτε κατακόρυφες για τον εντοπισμό των στηλών) την θέση και το μέγεθος των οποίων βελτιστοποιεί με τον γενετικό αλγόριθμο. Και η δύο παραπάνω μέθοδοι δημιουργούν ολιστικά το πλέγμα το οποίο απομονώνει την κάθε κηλίδα στην εικόνα.

3.2.4. Προσέγγιση Κηλίδα με κηλίδα

Σε αντίθεση με την ολιστική προσέγγιση η προσέγγιση κηλίδα με κηλίδα δημιουργεί το πλέγμα και απομονώνουν την κάθε κηλίδα ξεχωριστά, αφού πρώτα εντοπίσουν τα κέντρα όλων των κηλίδων της εικόνας. Ο εντοπισμός όλων των κηλίδων έχει δύο βασικά μειονεκτήματα. Αφενός απαιτούν πολύ μεγαλύτερη υπολογιστική ισχύς και είναι χρονοβόρα διαδικασία σε σχέση με μια ολιστική προσέγγιση, αφετέρου πρέπει να αντιμετωπιστεί το γεγονός ότι υπάρχουν πολλές κηλίδες σε μια εικόνα μικροσυστοιχιών που προέρχονται από ανιχνευτές που δεν έχουν υβριδοποιηθεί με τα δείγματα και ως εκ τούτου έχουν πολύ χαμηλή φωτεινότητα. Ωστόσο, η μέθοδοι που υλοποιούν μια προσέγγιση κηλίδα με κηλίδα έχουν μεγαλύτερη ακρίβεια στον εντοπισμό των κέντρων

των κηλίδων, και στην αντιστοίχιση μιας περιοχής γύρω από αυτές, αφού αντιμετωπίζουν το πρόβλημα των μικρών μετατοπίσεων των κηλίδων από τις συμμετρικές θέσεις.

Για την υλοποίηση μιας προσέγγισης κηλίδα με κηλίδα αρχικά εντοπίζονται οι φωτεινές κηλίδες της εικόνας με απλές μεθόδους κατάτμησης εικόνας. Σκοπός αυτής της κατάτμησης δεν είναι να βρεθούν ακριβώς τα εικονοστοιχεία σήματος και υποβάθρου, αλλά απλώς να βρεθούν όλα τα φωτεινά αντικείμενα της εικόνας, τα οποία είναι πιθανότατα οι φωτεινές κηλίδες. Έτσι, στο σημείο αυτό δεν χρησιμοποιούνται πολύπλοκες μέθοδοι κατάτμησης, ώστε να μην χρειαστούν πολλοί πόροι από το σύστημα. Ο εντοπισμός και διαχωρισμός των εικονοστοιχείων σήματος και υποβάθρου γίνεται αναλυτικά στο επόμενο στάδιο της επεξεργασίας της εικόνας για κάθε μια κηλίδα ξεχωριστά. Πολλές μέθοδοι για το εντοπισμό των φωτεινών κηλίδων περιορίζονται σε μια απλή κατωφλίωση (thresholding) της εικόνας [110-111]. Η εικόνα μετατρέπεται σε δυαδική όπου με 0 σημειώνονται τα εικονοστοιχεία χαμηλής φωτεινότητας και με 1 τα εικονοστοιχεία που μετέχουν σε φωτεινά αντικείμενα της εικόνας. Το κατώφλι πολλές φορές επιλέγεται ευριστικά [110] ενώ άλλες φορές χρησιμοποιούνται μέθοδοι μεταβλητού κατωφλιού [111] (adaptive thresholding) όπως είναι η μέθοδος του Ότσου [112]. Μια επίσης διαδεδομένη τεχνική για τον εντοπισμό των φωτεινών κηλίδων είναι η τεχνική Σύμπτωσης Προτύπων (Template Matching) [113]. Οι μέθοδοι που χρησιμοποιούν αυτήν την τεχνική [92, 114-115] εκμεταλλεύονται το γεγονός ότι τυπικά μια κηλίδα πρέπει να έχει κυκλικό σχήμα και η κατανομή της έντασης της να προσεγγίζει την κανονική κατανομή. Με αυτό τον τρόπο, γίνεται και μια επιλογή των φωτεινών αντικειμένων απορρίπτοντας τα φωτεινά αντικείμενα τα οποία δεν πληρούν αυτές τις προϋποθέσεις ως προς το σχήμα τους και την κατανομή της φωτεινότητας τους. Ωστόσο, εφαρμόζοντας την τεχνική σύμπτωσης προτύπων θα απορριφθούν και κηλίδες των οποίων το σχήμα αποκλίνουν πολύ από το κυκλικό λόγο ανομοιομορφίας της υβριδοποίησης, γεγονός όμως το οποίο δεν έχει πολλές συνέπειες σε αυτήν την φάση της επεξεργασίας διότι αυτές θα εντοπιστούν από την επόμενη διαδικασία μαζί με τις κηλίδες χαμηλής φωτεινότητας. Είναι προτιμότερο σε αυτήν την φάση να εντοπιστούν μόνο τα αντικείμενα που σίγουρα προέρχονται από την υβριδοποίηση των ανιχνευτών, χωρίς να συμπεριληφθούν και πολλά τεχνουργήματα σε αυτά, πληρώνοντας ένα τίμημα σε μη εντοπισμένες πραγματικές κηλίδες. Παρόμοια τεχνική με την σύμπτωσης προτύπων χρησιμοποιούν στην μέθοδό τους

και οι Brandle *et al.* [85] θεωρώντας και αυτοί ότι η κατανομή της έντασης μιας τυπική κηλίδα ακολουθεί κανονική κατανομή.

Μετά τον εντοπισμό των φωτεινών κηλίδων οι μέθοδοι που κάνουν χρήση της προσέγγισης κηλίδα με κηλίδα καλούνται να προσδιορίσουν τις θέσεις των κέντρων των κηλίδων χαμηλής φωτεινότητας. Αυτό γίνεται χρησιμοποιώντας τις συμμετρίες που έχει μια εικόνα μικροσυστοιχιών [114-117]. Συγκριμένα, σκοπός είναι να εξαχθούν διάφορα χαρακτηριστικά από το σύνολο των ήδη εντοπισμένων φωτεινών κηλίδων, με σκοπό να προσδιοριστούν με αρκετή ακρίβεια οι θέσεις της εικόνας όπου αναμένεται να υπάρχουν μη-υβριδοποιημένοι ανιχνευτές. Έτσι από τις εντοπισμένες φωτεινές κηλίδες υπολογίζονται οι μέσες τιμές διάφορων μεγεθών, όπως για παράδειγμα οι διάμετροι τους, η απόσταση μεταξύ δύο γειτονικών κηλίδων κ.α.. Ανατρέχοντας σε κάθε εντοπισμένη φωτεινή κηλίδα και χρησιμοποιώντας τις ανωτέρω τιμές μπορούν να εξαχθούν κατά προσέγγιση οι θέσεις γειτονικών κηλίδων χαμηλής φωτεινότητας. Η διαδικασία αυτή γίνεται επαναληπτικά για κάθε μια εντοπισμένη φωτεινή κηλίδα με σκοπό να προσδιοριστούν οι θέσεις κάθε μιας κηλίδας χαμηλής φωτεινότητας. Για τον λόγο αυτό και η συγκεκριμένη προσέγγιση αναφέρεται και ως προσέγγιση κηλίδα με κηλίδα. Πιο εξελιγμένες μέθοδοι αυτής της κατηγορίας κάνουν χρήση γράφων (graph models) [110, 118]. Η μέθοδος των Jung *et al.* είναι η κυριότερη μέθοδος, αυτής της κατηγορίας και είναι βασισμένη στον γράφο του K-στού κοντινότερου γείτονα (k-Nearest Neighbors graph) [119]. Για κάθε εντοπισμένη φωτεινή κηλίδα βρίσκεται η κοντινότερη γειτονική κηλίδα. Αν σε ένα ζευγάρι γειτονικών κηλίδων ισχύει ότι η κάθε κηλίδα από τις δύο έχει για κοντινότερη την άλλη κηλίδα του ζευγαριού, τότε αυτές αποτελούν ένα αμοιβαίο (reciprocal) ζευγάρι. Οι κηλίδες οι οποίες επαληθεύουν αυτήν την συμμετρία παραμένουν συνδεδεμένες. Στην συνέχεια εντοπίζονται οι περιοχές μέσα στην κάθε συστάδα όπου υπάρχουν συνδεδεμένες κηλίδες. Ο αλγόριθμος διατρέχει αυτά τα ζευγάρια και με διάφορους κανόνες εντοπίζει τις θέσεις των κηλίδων, οι οποίες δεν έχουν εντοπιστεί από την αρχική κατάτμηση, λόγω της χαμηλής τους φωτεινότητας.

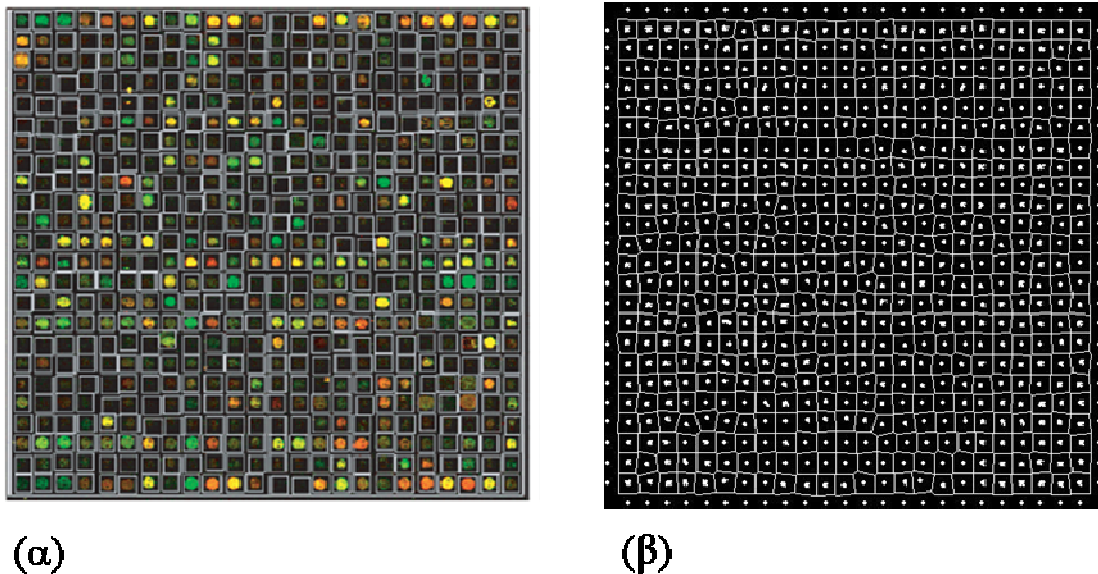
3.2.5. Τοποθέτηση Πλέγματος

Η τοποθέτηση του πλέγματος αφορά στην απομόνωση της κάθε κηλίδας σε ένα κελί, δηλαδή σε μια μικρή περιοχή της εικόνας. Αυτός είναι και ο σκοπός που εξυπηρετεί το παρόν στάδιο της επεξεργασίας μια εικόνας μικροσυστοιχιών. Το στάδιο της κατάτμησης, το οποίο ακολουθεί, εφαρμόζεται σε κάθε ένα κελί ξεχωριστά, γεγονός που απαιτεί η τοποθέτηση των κηλίδων στα κελία τους να είναι ακριβής. Μέσα στο κάθε κελί πρέπει αφενός να περιέχονται όλα τα εικονοστοιχεία σήματος της κηλίδας, και αφετέρου να μην υπάρχουν εικονοστοιχεία από γειτονικές κηλίδες.

Οι ολιστικές μέθοδοι εξορισμού δημιουργούν απευθείας το πλέγμα που θα διαχωρίσει τις κηλίδες της εικόνας. Κάποιες από αυτές χτίζουν εξαρχής το πλέγμα στην σωστή του θέση [104-109], κάποιες άλλες όμως [75, 97-98] δημιουργούν ένα πλέγμα χρησιμοποιώντας τις παραμέτρους που αναφέρθηκαν στο εδάφιο των ολιστικών προσεγγίσεων και στην συνέχεια το τοποθετούν στην θέση που πρέπει, είτε χειροκίνητα [75], είτε με αυτόματο τρόπο [97-98]. Το γεγονός ότι κάποιες από τις ολιστικές μεθόδους λοιπόν απαιτούν την τοποθέτηση του παραγόμενου πλέγματος αποτελεί και τον λόγο για τον οποίο το βήμα αυτό δεν συμπεριλήφθηκε στον κλάδο του σχήματος 3.2 για την προσέγγιση κηλίδα με κηλίδα. Ωστόσο, οι μέθοδοι οι οποίες κατεξοχήν χρειάζονται αυτό το βήμα είναι αυτές που ακολουθούν την προσέγγιση κηλίδα με κηλίδα.

Όσον αφορά στην προσέγγιση κηλίδας με κηλίδα, οι περισσότερες από τις μεθόδους [110, 117-119] δεν εφαρμόζουν κάποια ευφυή τεχνική για την δημιουργία των κελιών. Έχοντας εξάγει χαρακτηριστικά για το μέγεθος της κάθε κηλίδας απλώς την καταχωρούν σε ένα τετράγωνο κελί κεντροθετημένο στο εντοπισμένο κέντρο της κηλίδας και το οποίο έχει μέγεθος πλευράς ίσο με την απόσταση δύο γειτονικών κηλίδων. Η προσέγγιση όμως αυτή μπορεί να προκαλέσει προβλήματα λόγω των μικρών μετατοπίσεων των κηλίδων, αλλά και του ενδεχόμενου τα κέντρα των κηλίδων να μην έχουν εντοπιστεί με ακρίβεια. Σε αυτήν την περίπτωση είναι πολύ πιθανό το κελί είτε να μην περιέχει όλα τα εικονοστοιχεία σήματος της κηλίδας, είτε να περιέχει εικονοστοιχεία σήματος γειτονικών κηλίδων. Για τον λόγο αυτό μερικές μέθοδοι [92, 114-115] κάνουν χρήση του Βορονόι διαγράμματος [120], το οποίο όπως θα δούμε και στο επόμενο κεφάλαιο αποδίδει την χωρικά βέλτιστη

περιοχή σε κάθε κελί, συνυπολογίζοντας τα κέντρα και των γειτονικών κηλίδων. Στο σχήμα 3.8 φαίνεται το αποτέλεσμα της μεθόδου των Jung και Gho, οι οποίοι απέδωσαν ένα σταθερό τετράγωνο κελί σε κάθε κηλίδα, αλλά και το αποτέλεσμα ενός πλέγματος που παρήχθει με την χρήση Βορονόι διαγράμματος.



Σχήμα 3.8: Η τοποθέτηση πλέγματος σε μια εικόνα μικροσυστοιχιών. (α) Τοποθέτηση σταθερού κελιού σε κάθε μία κηλίδα, (β) εφαρμογή του διαγράμματος Βορονόι.

Το Βορονόι διάγραμμα χρησιμοποιήθηκε και από την προτεινόμενη μεθοδολογία που αναπτύχθηκε στα πλαίσια της παρούσας διατριβής. Τα πλεονεκτήματα του Βορονόι διαγράμματος θα μελετηθούν εκτενώς σε σχέση με την τετριμμένη προσέγγιση των σταθερών τετράγωνων κελιών, δεδομένου ότι η προτεινόμενη μεθοδολογία για την τον εντοπισμό των κηλίδων και την τοποθέτηση πλέγματος μπορεί να επεξεργάζεται και εικόνες στις οποίες οι κηλίδες είναι χωροθετημένες σε εξαγωνικό πλέγμα.

3.2.6.Βελτίωση του πλέγματος

Ως τελευταίο βήμα των μεθόδων για τον εντοπισμό των κηλίδων και την τοποθέτηση του πλέγματος, πολλές μέθοδοι [89, 97, 114-115, 121-122] επιχειρούν να βελτιώσουν το πλέγμα το οποίο έχει παραχθεί. Η διαδικασία αυτή έχει μεγάλη σημασία κυρίως για τις μεθόδους ολιστικής προσέγγισης δεδομένου ότι οι μέθοδοι αυτές δεν μπορούν να

αντιμετωπίσουν τις μικρές μετατοπίσεις των κηλίδων από τις συμμετρικές σχέσεις και γενικά τα τοπικά προβλήματα τα οποία ενδέχεται να παρουσιαστούν σε μία εικόνα. Στην προσπάθεια τους αυτή επιστρατεύουν διάφορους μετασχηματισμούς, οι οποίοι δέχονται το παραγόμενο πλέγμα ως είσοδο. Οι κυριότεροι μετασχηματισμοί που έχουν χρησιμοποιηθεί για τον σκοπό αυτό είναι: (α) ο μετασχηματισμός Affine [123] τον οποίο χρησιμοποίησε ο Galinsky [114-115], (β) ο μετασχηματισμός Hough [124] που χρησιμοποίησαν οι Bozinov *et al.* [89] και τέλος (γ) ο μετασχηματισμός Radon [125] που πρώτοι εφήρμοσαν οι Antoniol *et al.* [121].

Μετά την ολοκλήρωση του σταδίου του εντοπισμού των κηλίδων και της απομόνωσης κάθε κηλίδας στην εικόνα σε ένα κελί, ακολουθεί η κατάτμηση των κηλίδων η οποία γίνεται σε κάθε ένα κελί ξεχωριστά.

3.3. Κατάτμηση των κηλίδων

Όπως αναφέρθηκε και ανωτέρω η κατάτμηση μιας εικόνας έχει σκοπό την ομαδοποίηση εικονοστοιχείων με παρόμοια χαρακτηριστικά.. Με αυτό τον τρόπο η διαδικασία της κατάτμησης διαχωρίζει την εικόνα σε περιοχές ή εντοπίζει αντικείμενα. Είναι πάρα πολλές οι εφαρμογές για τις οποίες απαραίτητη προϋπόθεση για περαιτέρω επεξεργασία είναι ο προσδιορισμός περιοχών η αντικειμένων μέσα σε μια εικόνα. Ένα παράδειγμα τέτοιων εφαρμογών είναι και η επεξεργασία των εικόνων των μικροσυστοιχιών. Απαραίτητη προϋπόθεση για την ποσοτικοποίηση της έντασης της κάθε κηλίδας είναι ο εντοπισμός της περιοχής σήματος της κηλίδας από την οποία θα υπολογίσουμε την ένταση της. Έτσι η κατάτμηση στην επεξεργασία των εικόνων των μικροσυστοιχιών εμπίπτει στον διαχωρισμό των εικονοστοιχείων σήματος από τα εικονοστοιχεία υποβάθρου.

Έρευνες στο πεδίο των μικροσυστοιχιών έχουν δείξει ότι η κατάτμηση των εικόνων παίζει καθοριστικό ρόλο στην ποιότητα των εξαγόμενων δεδομένων [126]. Μια αποτυχία στο στάδιο της κατάτμησης μπορεί να προκαλέσει λανθασμένα συμπεράσματα για κάποια βιολογική λειτουργία. Πολλές είναι η εργασίες οι οποίες έχουν παρουσιαστεί στην βιβλιογραφία για την κατάτμηση των εικόνων των μικροσυστοιχιών. Οι μέθοδοι που

προτείνονται στις εργασίες αυτές μπορούν να κατηγοριοποιηθούν σε τέσσερις βασικές κατηγορίες [81]:

- Μέθοδοι σταθερού ή μεταβλητού κύκλου (Fixed and adaptive circle).
- Μέθοδοι βασισμένες στο ιστόγραμμα της εικόνας (histogram-based).
- Μέθοδοι μεταβλητού σχήματος (adaptive shape).
- Μέθοδοι βασισμένες σε τεχνικές ομαδοποίησης (clustering-based).

Γενικεύοντας περαιτέρω την παραπάνω κατηγοριοποίηση των μεθόδων μπορούμε να αναφέρουμε ότι οι πρώτες τρεις κατηγορίες εμπίπτουν σε μεθόδους βασισμένες σε τεχνικές επεξεργασία εικόνας (Image Processing) ενώ η τέταρτη είναι βασισμένη σε τεχνικές μηχανικής μάθησης (Machine Learning). Σύμφωνα με το γεγονός αυτό το σχήμα 3.9 παρουσιάζει μια γενική επισκόπηση των μεθόδων για την κατάτμηση των εικόνων των μικροσυστοιχιών με την μορφή δέντρου.

Σε αυτά τα πλαίσια κατηγοριοποίησης των μεθόδων κατάτμησης η παρούσα διατριβή προτείνει την χρήση τεχνικών ταξινόμησης για την κατάτμηση των εικόνων. Όπως φαίνεται και στο σχήμα 3.9 η προσέγγιση αυτή μπορεί να καταταγεί στις μεθόδους κατάτμησης με χρήση τεχνικών μηχανικής μάθησης. Όπως θα δούμε και παρακάτω οι τεχνικές αυτές έχουν το πλεονέκτημα ότι ταξινομούν άμεσα τα εικονοστοιχεία τις εικόνας σε εικονοστοιχεία σήματος ή υποβάθρου. Σε αντίθεση οι τεχνικές ομαδοποίησης δημιουργούν ομάδες εικονοστοιχείων, οι οποίες πρέπει να χαρακτηριστούν ως σήμα ή υπόβαθρο σύμφωνα με μια σειρά κανόνων. Έτσι το κάθε εικονοστοιχείο χαρακτηρίζεται έμμεσα από τον χαρακτηρισμό της ομάδας στην οποία ανήκει. Στα παρακάτω εδάφια θα αναπτυχθούν οι μεθοδολογίες που έχουν χρησιμοποιηθεί για την κατάτμηση των κηλίδων σύμφωνα με το δενδρόγραμμα του σχήματος 3.9.

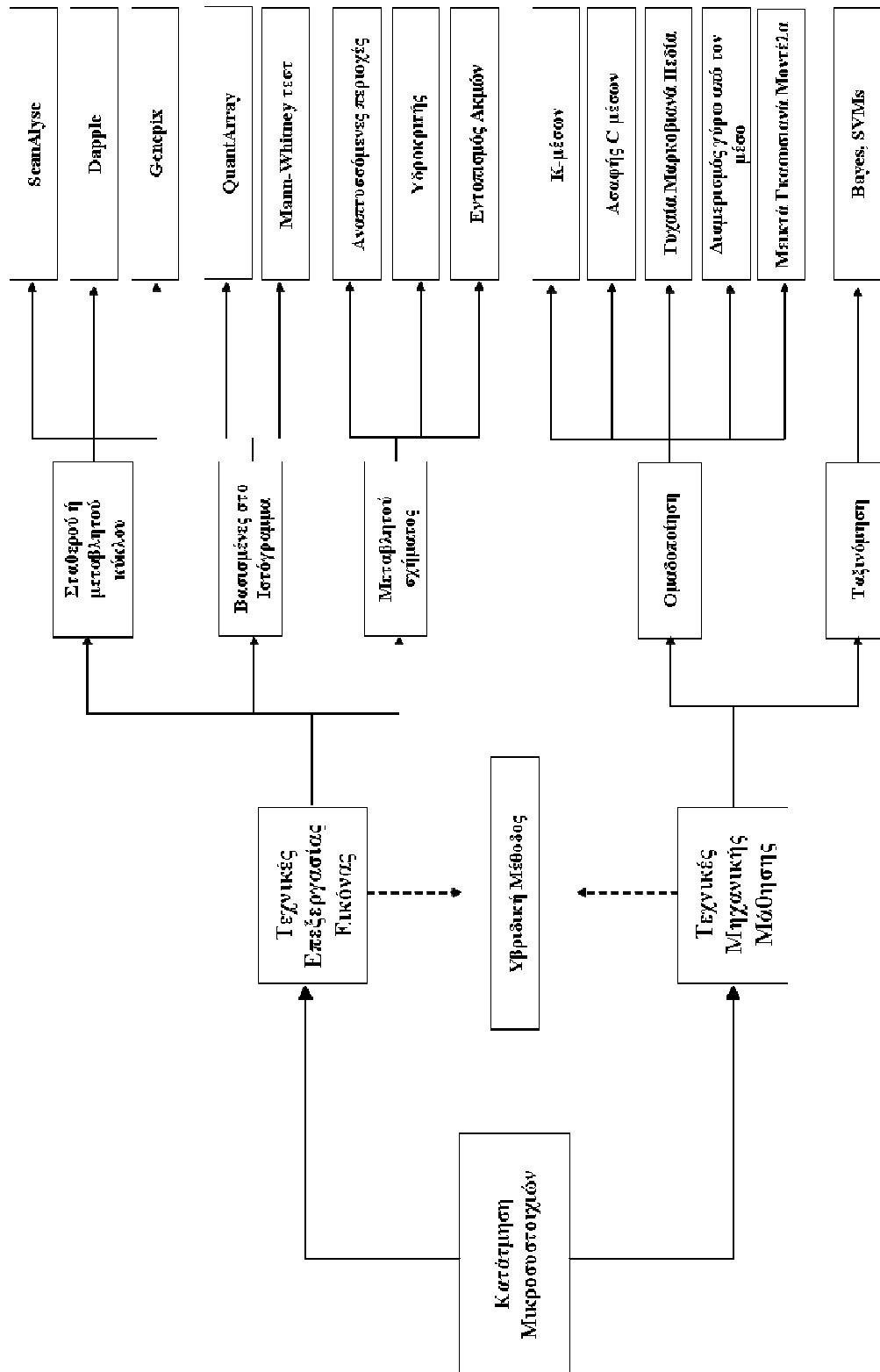
3.3.1. Μέθοδοι βασισμένες σε τεχνικές επεξεργασίας εικόνων

Οι μέθοδοι, οι οποίες είναι βασισμένες σε τεχνικές επεξεργασίας εικόνας χωρίζονται σε τρεις κατηγορίες όπως φαίνεται στο δενδρόγραμμα του σχήματος 3.9. Οι μέθοδοι σταθερού ή μεταβλητού κύκλου είναι μέθοδοι, οι οποίες αναπτύχθηκαν κυρίως στα πλαίσια εμπορικών πακέτων για την επεξεργασία των εικόνων. Οι μέθοδοι αυτές απλώς υπερθέτουν ένα κύκλο επάνω από κάθε κηλίδα της μικροσυστοιχίας, Τα εικονοστοιχεία τα

οποία βρίσκονται εντός του κύκλου θεωρούνται εικονοστοιχεία σήματος ενώ τα εικονοστοιχεία εκτός του κύκλου θεωρούνται εικονοστοιχεία υποβάθρου. Οι μέθοδοι οι οποίες είναι βασισμένες στο ιστόγραμμα τις εικόνας, υπολογίζουν ένα κατώφλι φωτεινότητας πάνω από το οποίο ένα εικονοστοιχείο θεωρείται εικονοστοιχείο σήματος ενώ κάτω από το οποίο θεωρείται εικονοστοιχείο υποβάθρου. Με την πάροδο του χρόνου πιο εξελιγμένες τεχνικές επεξεργασίας εικόνας άρχισαν να χρησιμοποιούνται για την κατάτμηση των εικόνων των μικροσυστοιχιών. Οι μέθοδοι αυτές ήταν κυρίως βασισμένες στους αλγορίθμους του υδροκριτή (Watershed) και των αναπτυσσόμενων περιοχών (Seed Region Growing - SRG), καθώς επίσης και στον εντοπισμό των ακμών (edge detection) των κηλίδων.

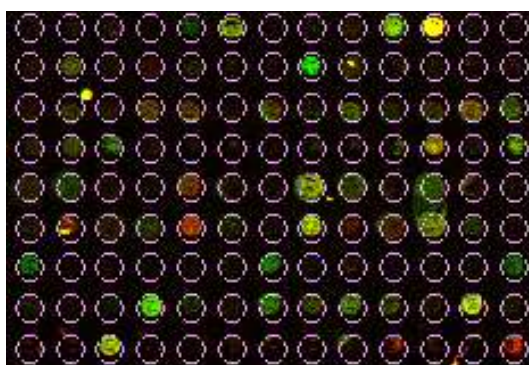
3.3.1.1.Μέθοδοι σταθερού ή μεταβλητού κύκλου

Η προσέγγιση του σταθερού κύκλου (fixed circle) αποτελεί την πρώτη προσέγγιση μεθόδων για κατάτμηση εικόνων μικροσυστοιχιών. Η προσέγγιση αυτή χρησιμοποιήθηκε για πρώτη φορά από το πακέτο ScanAlyze [75], το οποίο δημιουργήθηκε από τον Eisen το 1999 στο Πανεπιστήμιο του Στάνφορντ, αλλά εφαρμόστηκε και αργότερα από άλλες ερευνητικές ομάδες [127]. Η μέθοδος σταθερού κύκλου που χρησιμοποιείται για την κατάτμηση της εικόνας απλώς τοποθετεί ένα κύκλο γύρω από κάθε κηλίδα. Η θέση και η διάμετρος του κύκλου που εφαρμόζεται μπορεί να προσαρμοστεί από τον χρήστη, η διάμετρος όμως που ορίζει ο χρήστης είναι η ίδια για όλες τις κηλίδες της εικόνας. Όλα τα εικονοστοιχεία τα



Σχήμα 3.9: Δενδρόγραμμα των μεθόδων που έχουν παρουσιαστεί στην βιβλιογραφία για την κατάτμηση των εικόνων των μικροστοιχιών.

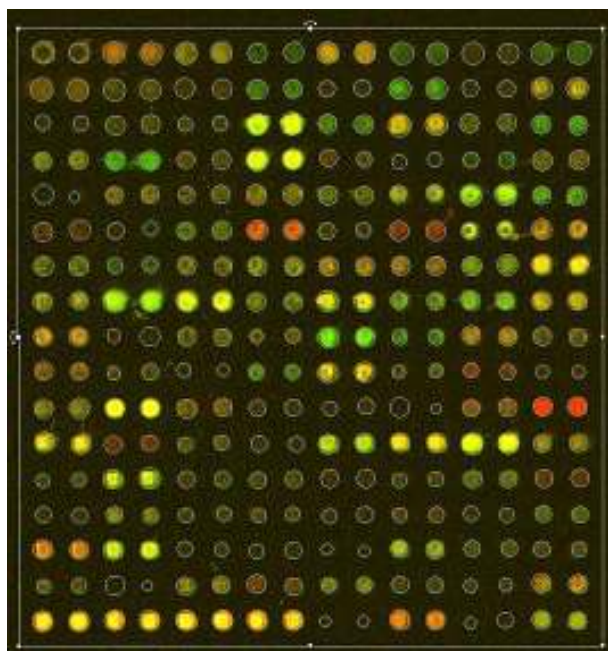
οποία βρίσκονται μέσα στον κύκλο θεωρούνται εικονοστοιχεία σήματος, ενώ όλα τα εικονοστοιχεία εκτός του κύκλου, αλλά μέσα στο κελί της εικόνας που περιέχει μια κηλίδα (όπως ορίστηκε από την διαδικασία της τοποθέτησης του πλέγματος), θεωρούνται εικονοστοιχεία υποβάθρου. Γίνεται φανερό ότι παρά το γεγονός ότι η μέθοδος αυτή είναι απλή στην υλοποίηση της, έχει πάρα πολλές αδυναμίες. Αφενός είναι απαραίτητη η παρέμβαση του χρήστη, και αφετέρου αδυνατεί να κάνει κατάτμηση σε μη ιδανικές εικόνες μικροσυστοιχιών, όπως αυτές αναφέρθηκαν στο πρώτο εδάφιο του κεφαλαίου αυτού. Στο σχήμα 3.10 φαίνεται το αποτέλεσμα της εφαρμογής μιας μεθόδου σταθερού κύκλου σε μια περιοχή μια εικόνας μικροσυστοιχιών.



Σχήμα 3.10: Κατάτμηση με μέθοδο σταθερού κύκλου.

Η άρση του περιορισμού της σταθερής ακτίνας του κύκλου για όλες τις κηλίδες της εικόνας ήταν το αμέσως επόμενο βήμα για τις μεθόδους επεξεργασίας εικόνων μικροσυστοιχιών. Έτσι αναπτύχθηκαν οι μέθοδοι μεταβλητού κύκλου [128] οι οποίες όμως και αυτές διατηρούσαν την θεώρηση ότι το σχήμα των κηλίδων είναι πάντα κυκλικό. Κάποιες από τις μεθόδους αυτές απαιτούσαν την παρέμβαση του χρήστη όσον αφορά στην ακτίνα της κάθε κηλίδας, ενώ άλλες προσδιορίζουν την ακτίνα της κάθε κηλίδας με αυτόματο τρόπο. Για παράδειγμα το πρόγραμμα Dapple [129] υπολογίζει την ακτίνα του κύκλου που θα υπερτεθεί πάνω από μια κηλίδα κάνοντας χρήση της Λαπλασιανής τεχνικής εντοπισμού ακμών (Laplacian-based edge detection). Αντιθέτως, οι μέθοδοι οι οποίες απαιτούν από τον χρήστη να ορίσει το μέγεθος της κάθε κηλίδας είναι εξαιρετικά χρονοβόρες λόγω του μεγάλου αριθμού των κηλίδων που βρίσκονται σε κάθε εικόνα. Μια τέτοια μέθοδο χρησιμοποιούσαν οι πρώτες εκδόσεις του πακέτου Genepix [130]. Στο σχήμα 3.11 παρουσιάζεται η περιοχή μιας εικόνας, οι οποία έχει επεξεργαστεί με το πρόγραμμα Genepix. Είναι εμφανές ότι το αποτέλεσμα του προγράμματος Genepix προσεγγίζει με καλύτερο τρόπο την εικόνα μικροσυστοιχιών από το πρόγραμμα ScanAlyze. Τα γονίδια τα

οποία έχουν μερικώς υβριδοποιηθεί και έχουν δημιουργήσει μικρότερες κηλίδες στην εικόνα έχουν προσομοιωθεί με έναν κύκλο ο οποίος αντιστοιχεί στο μέγεθος τους.



Σχήμα 3.11: Κατάτμηση με το πρόγραμμα Genepix

Τέλος πρέπει να αναφέρουμε την μέθοδο των Rueda και Qin [131], οποίοι επιχείρησαν να κάνουν ένα ακόμα βήμα γενίκευσης των μεθόδων σταθερού ή μεταβλητού κύκλου, αίροντας τον περιορισμό του κυκλικού σχήματος της κηλίδας, και προσπαθώντας να προσδιορίσουν την βέλτιστη έλλειψη η οποία προσομοιώνει καλύτερα την κάθε κηλίδα της εικόνας.

3.3.1.2. Μέθοδοι βασισμένες στο ιστόγραμμα της εικόνας

Με την πάροδο του χρόνου οι μέθοδοι κατάτμηση απέβαλαν ολοένα τους περιορισμούς για το μέγεθος των κηλίδων, το σχήμα των κηλίδων και τα άλλα χαρακτηριστικά μίας ιδανικής εικόνας μικροσυστοιχιών, κάνοντας χρήση των φωτεινότητων των εικονοστοιχείων και του ιστογράμματος της εικόνας, με σκοπό να διαχωρίσουν τα εικονοστοιχεία σήματος και υποβάθρου. Πιο συγκεκριμένα οι μέθοδοι αυτές προσπαθούν μέσα στο κελί της κάθε κηλίδας να υπολογίσουν μια τιμή κατώφλιού στην φωτεινότητα των εικονοστοιχείων. Όσα εικονοστοιχεία έχουν φωτεινότητα μεγαλύτερη από αυτό το κατώφλι αυτόματα χαρακτηρίζονται ως εικονοστοιχεία σήματος, ενώ αντίθετα όσα έχουν

φωτεινότητα χαμηλότερη από το κατώφλι χαρακτηρίζονται ως εικονοστοιχεία υποβάθρου [97,132-136].

Ο πιο διαδεδομένος τρόπος προσδιορισμού του κατωφλιού στο πεδίο της κατάτμησης εικόνων μικροσυστοιχιών [132-134] είναι βασισμένος στο τεστ των Mann και Whitney [137]. Το Mann-Whitney τεστ, το οποίο λέγεται και Wilcoxon τεστ, είναι ένα στατιστικό τεστ, το οποίο στοχεύει στην εύρεση μιας σημαντικής διαφοράς μεταξύ των πληθυσμών δύο συνόλων A και B ανεξαρτήτως της κατανομής που ακολουθούν τα στοιχεία των δύο πληθυσμών. Στην περίπτωση των εικόνων των μικροσυστοιχιών οι δύο πληθυσμοί είναι οι φωτεινότητες εικονοστοιχείων σήματος και υποβάθρου. Ο αρχικός αυτός διαχωρισμός των εικονοστοιχείων έχει γίνει με μια κυκλική μάσκα όπως ακριβώς διαχωρίζονται τα εικονοστοιχεία στις μεθόδους σταθερού κύκλου. Το τεστ είναι βασισμένο στο Wilcoxon άθροισμα των τάξεων. Αυτό ορίζεται ως το άθροισμα της τάξης των στοιχείων ενός πληθυσμού εικονοστοιχείων A , τα οποία έχουν ταξινομηθεί μαζί με τα εικονοστοιχεία στοιχείων του άλλου πληθυσμού B . Τα εικονοστοιχεία δηλαδή και των δύο πληθυσμών έχουν τεθεί σε αύξουσα σειρά ανάλογα με την τιμή της φωτεινότητας τους. Το εικονοστοιχεία με την μικρότερη φωτεινότητα και από τους δύο πληθυσμούς θεωρείται ότι έχει τάξη 1, ενώ το μεγαλύτερο τάξη $n_a + n_b$, εάν υποθέσουμε ότι το σύνολο A έχει n_a στοιχεία και το σύνολο B έχει n_b . Το Wilcoxon άθροισμα για το σύνολο A θα είναι:

$$SW_A = rank(a_1) + rank(a_2) + \dots + rank(a_\mu), \quad (3.1)$$

Στην συνέχεια υπολογίζονται τα στατιστικά μεγέθη U_β και U_a αντίστοιχα για τα δύο σύνολα B και A :

$$U_\beta = n_b \cdot n_a + \frac{n_b \cdot (n_b - 1)}{2} - SW_\beta, \quad (3.2)$$

και

$$U_a = n_b \cdot n_a + \frac{n_a \cdot (n_a - 1)}{2} - SW_a, \quad (3.3)$$

Έπειτα η τιμή U_i που προκύπτει από το σύνολο με τα λιγότερα στοιχεία συγκρίνεται με μία τιμή U η οποία προκύπτει από ένα πίνακα πιθανοτήτων του τεστ και έχει να κάνει και με τον αριθμό των στοιχείων των δύο πληθυσμών. Αν τα δύο σύνολα έχουν τον ίδιο αριθμό στοιχείων τότε προς σύγκριση με την τιμή U τίθεται η μικρότερη από τις τιμές U_β και U_α . Έως ότου οι τιμή του U_i να συγκλίνει στην τιμή U , οι διαδικασίες συνεχίζονται επαναληπτικά μεταφέροντας εικονοστοιχεία από τον ένα πληθυσμό στον άλλο.

Οι μέθοδοι ιστογράμματος ενδέχεται να μην είναι τόσο σταθερές όταν ένα κελί περιλαμβάνει εικονοστοιχεία από γειτονικές κηλίδες ή τεχνουργήματα. Επίσης τα αποτελέσματα αυτών των μεθόδων είναι απογοητευτικά όταν η φωτεινότητα της κηλίδας είναι χαμηλή λόγω περιορισμένης υβριδοποίησης, και άρα ο διαχωρισμός των εικονοστοιχείων σήματος από τα εικονοστοιχεία υποβάθρου γίνεται πολύ δύσκολο.

3.3.1.3. Μέθοδοι Μεταβλητού σχήματος

Η κατηγορία αυτή αναφέρεται στις μεθόδους, οι οποίες είναι βασισμένες σε εξελιγμένους αλγόριθμους επεξεργασίας εικόνας. Το βασικό πλεονέκτημα αυτών των προσεγγίσεων αφορά στο γεγονός ότι είναι απαλλαγμένες από κάθε είδους προϋποθέσεις και περιορισμούς σχετικά με το σχήμα και το μέγεθος των κηλίδων. Οι αλγόριθμοι επεξεργασίας εικόνας που κατά κύριο λόγο έχουν χρησιμοποιηθεί για την κατάτμηση των κηλίδων είναι ο αλγόριθμος των αναπτυσσόμενων περιοχών και ο αλγόριθμος του υδροκριτή, ενώ ορισμένες από τις μεθόδους εντοπίζουν τις ακμές των κηλίδων.

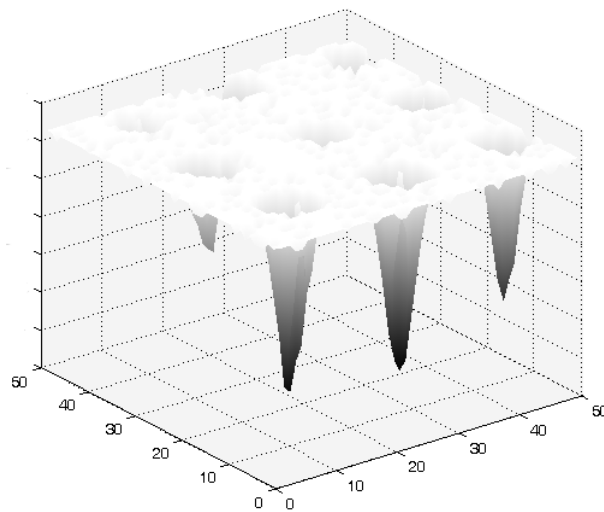
Ο αλγόριθμος των αναπτυσσόμενων περιοχών (Seed Region Growing – SRG) [138] επιλέγει τυχαία μικρές περιοχές από εικονοστοιχεία, που ονομάζονται «σπόροι» (seeds), και οι οποίοι χρησιμοποιούνται ως σημεία από τα οποία ξεκινά η ανάπτυξη των περιοχών. Στις εικόνες μικροσυστοιχιών «σπόροι» ορίζονται όχι μόνο για να αναπτυχθούν οι περιοχές του σήματος, αλλά και για να αναπτυχθούν οι περιοχές του υπόβαθρου. Σε κάθε επανάληψη ο αλγόριθμος συλλέγει όλα τα εικονοστοιχεία τα οποία συνορεύουν με την περιοχή «σπόρο» και τα ιεραρχεί σύμφωνα με διάφορα κριτήρια. Το πιο διαδεδομένο

κριτήριο [139] εξαρτάται μόνο από την φωτεινότητα των εικονοστοιχείων και δίνεται από την σχέση:

$$Cr(i) = |I(i) - \overline{I(j)}|, \quad j \in S, \quad (3.4)$$

όπου S είναι η αναπτυσσόμενη περιοχή, $I(i)$ είναι η φωτεινότητα του εικονοστοιχείου i το οποίο είναι ένα εικονοστοιχείου γειτονικό της περιοχής S , ενώ, $\overline{I(j)}$ είναι η μέση τιμή της φωτεινότητας της αναπτυσσόμενης περιοχής S στην παρούσα επανάληψη του αλγορίθμου. Αφού ο αλγόριθμος ιεραρχήσει τα εικονοστοιχεία λαμβάνει το εικονοστοιχείο, το οποίο δίνει την μικρότερη τιμή του κριτηρίου Cr , και αποφασίζει για αυτό το εικονοστοιχείο αν πρέπει να ανείκει στην ομάδα που ορίζουν οι «σπόρων» (σήμα ή υπόβαθρο) ή αν είναι συνοριακό εικονοστοιχείο της περιοχής. Εάν όλα τα γειτονικά του εικονοστοιχεία ανήκουν στην ίδια περιοχή τότε και αυτό κατατάσσεται σε αυτήν την περιοχή. Σε αντίθετη περίπτωση το εικονοστοιχείο αυτό θεωρείται συνοριακό εικονοστοιχείο. Δύο διαφορετικές προσεγγίσεις της εφαρμογής του αλγορίθμου αναπτυσσόμενων περιοχών έχουν προταθεί για την κατάτμηση των κηλίδων. Η πρώτη προσέγγιση των Yang *et al.* [140] χρησιμοποιεί ως «σπόρους» για τα εικονοστοιχεία σήματος μια τετράγωνη περιοχή κεντροθετημένη στο εικονοστοιχείο με την υψηλότερη φωτεινότητα του κελιού και μέγεθος ίσο με το $1/3$ του κελιού. Ως «σπόρος» υποβάθρου χρησιμοποιούνται τα εικονοστοιχεία των ακμών του κελιού. Η δεύτερη προσέγγιση [141] χρησιμοποιεί ως «σπόρο» σήματος το κεντρικό εικονοστοιχείο του κελιού ενώ ως «σπόρο» υποβάθρου τις τέσσερις γωνίες του κελιού.

Ο δεύτερος αλγόριθμος που έχει χρησιμοποιηθεί από τις μεθόδους μεταβλητού σχήματος είναι ο αλγόριθμος του υδροκριτή (Watershed) [142]. Ο αλγόριθμος του υδροκριτή είναι ένας πολύ διαδεδομένος αλγόριθμος, ο οποίος είναι βασισμένος στην μαθηματική μορφολογία, και ο οποίος έχει χρησιμοποιηθεί πολύ και στο πεδίο της κατάτμησης εικόνων μικροσυστοιχείων. Η εικόνα θεωρείται ως μια ανάγλυφη τοποθεσία, η οποία έχει πολλά πηγάδια στα σημεία που υπάρχουν κηλίδες όπως δείχνει το σχήμα 3.12.



Σχήμα 3.12: Μέρος μιας εικόνας μικροσυστοιχιών θεωρούμενη ως ανάγλυφη τοποθεσία.

Ο αλγόριθμος θεωρεί ότι στην ανάγλυφη τοποθεσία βρέχει. Ανά πάσα στιγμή όσο το νερό γεμίζει την τοποθεσία εικόνα κατατέμενεται από τις προβολές της επιφάνειας του νερού. Κάποιες ερευνητικές ομάδες έχουν επιλέξει να υλοποιήσουν όλη την επεξεργασία των εικόνων των μικροσυστοιχιών με χρήση μαθηματική μορφολογίας [99-102]. Έτσι για την διαδικασία της κατάτμησης κάνουν χρήση του υδροκριτή. Σύμφωνα με την βιβλιογραφία ο υδροκριτής εφαρμόζεται είτε στην αρχική εικόνα [143-144], είτε στην κλίση (gradient) της εικόνας [145]. Η έξοδος του αλγορίθμου είναι η εικόνα χωρισμένη σε περιοχές, τις οποίες ορίζει η επιφάνεια του νερού.

Πέραν των δύο βασικών αλγορίθμων έχουν γίνει και άλλες μεμονωμένες προσπάθειες βασισμένες σε τεχνικές επεξεργασίας εικόνας. Οι περισσότερες από αυτές είναι βασισμένες σε τεχνικές εντοπισμού των ακμών των κηλίδων. Συγκεκριμένα, ο Kornaros [146] κάνει χρήση της τεχνικής του Sobel [147], ενώ οι Kim *et al.* [148] χρησιμοποιούν ένα Γκαουσιανό φίλτρο για να εντοπίσουν τις ακμές. Εντοπισμό των ακμών των κηλίδων έχουν προτείνει και οι Srinark και Kambhamettu [149], οι οποίοι κάνουν χρήση μοντέλων ενεργών περιγραμμάτων (Active Contours). Τέλος οι Gjerstad *et al.* [150-151], χρησιμοποίησαν μια μεθόδου βασισμένη στην τεχνική σύμπτωσης προτύπων, θεωρώντας ότι η κηλίδα έχει ένα τυπικό κυκλικό σχήμα και η κατανομή της φωτεινότητας της ακολουθεί δισδιάστατη κανονική κατανομή.

3.3.2.Μέθοδοι βασισμένες σε τεχνικές Μηχανικής Μάθησης

Οι τελευταίες μέθοδοι που εμφανίζονται στην βιβλιογραφία χρησιμοποιούν τεχνικές μηχανικής μάθησης και πιο συγκεκριμένα ομαδοποίησης. Γενικά η κατάτμηση μιας οποιοσδήποτε εικόνας, τα τελευταία έχει στραφεί προς τις μεθόδους ομαδοποίησης, οι οποίες εξυπηρετούν τον ίδιο ακριβώς σκοπό με αυτόν της κατάτμησης, αφού θεωρούν ως δείγματα τα χαρακτηριστικά των εικονοστοιχείων. Αλγόριθμοι όπως ο *K*-μέσων (*K*-means), ο ασαφής αλγόριθμος *C*-μέσων, ο αλγόριθμος Προσδοκίας-Μεγιστοποίησης (Expectation-Maximization-EM), ο αλγόριθμος διαμερισμού γύρω από το μέσο (Partitioning Around Medoid - PAM) και τα τυχαία Μαρκοβιανά πεδία (Markov Random Fields – MRF) έχουν χρησιμοποιηθεί με σκοπό να ομαδοποιήσουν εικονοστοιχεία της εικόνας με παρόμοια χαρακτηριστικά και να αποφανθούν αν είναι εικονοστοιχεία σήματος ή υποβάθρου. Τέλος είναι αξιοσημείωτη μια υβριδική μέθοδος που συνδυάζει τις μεθόδους που χρησιμοποιούν τεχνικές επεξεργασίας εικόνας με αυτές που χρησιμοποιούν τεχνικές μηχανικής μάθησης.

3.3.2.1.Κατάτμηση με τον αλγόριθμο *K*-μέσων

Η κατάτμηση αυτής της κατηγορίας είναι βασισμένη στο κλασικό αλγόριθμο ομαδοποίησης *K*-μέσων (*K*-means) [152]. Ο αλγόριθμος, τον οποίο θα περιγράψουμε πιο αναλυτικά στα επόμενα κεφάλαια, ξεκινά με την αρχικοποίηση των κεντροειδών (centroids) και στην συνέχεια επαναληπτικά κατατάσσει όλα τα εικονοστοιχεία της εικόνας σε μια από τις ομάδες. Η απόφαση λαμβάνεται σύμφωνα με την απόσταση που έχει το κάθε εικονοστοιχείο από το κεντροειδές της κάθε ομάδας. Στο πεδίο έρευνας των μικροσυστοιχιών ο αλγόριθμος *K*-μέσων έχει εφαρμοστεί από πολλές ερευνητικές ομάδες, με διάφορες παραλλαγές [153-160]. Οι περισσότερες μέθοδοι από αυτές τροφοδοτούν τον αλγόριθμο *K*-μέσων μόνο με τις φωτεινότητες των εικονοστοιχείων ως χαρακτηριστικό, και ορίζουν τον αριθμό των ομάδων ίσο με 2 (μία ομάδα για τα εικονοστοιχεία σήματος και μια ομάδα για τα εικονοστοιχεία υποβάθρου). Ωστόσο, κάποιες από αυτές τις μεθόδους καινοτομούν σε ορισμένα θέματα όπως είναι η εξαγωγή των χαρακτηριστικών και αριθμός των ομάδων. Οι Wu *et al.* [156] δημιούργησαν ένα διάλυμα χαρακτηριστικών για να τροφοδοτήσουν τον αλγόριθμο, το οποίο αποτελούνταν από τις φωτεινότητες των εικονοστοιχείων αλλά και από τις συντεταγμένες τους στην εικόνα.

Αξιοσημείωτη είναι και η εργασία των Wang *et al.* [159] οι οποίοι είναι και οι πρώτοι που χρησιμοποίησαν τον αλγόριθμο θέτοντας τον αριθμό των ομάδων ίσο με 3. Ωστόσο, στην εργασία τους αυτή δεν δόθηκε καμιά φυσική υπόσταση στην τρίτη ομάδα εικονοστοιχείων, την οποία απλώς την συγχώνευσαν σε ύστερο χρόνο με την ομάδα των εικονοστοιχείων σήματος. Αντίθετα, η μεθοδολογία η οποία προτείνεται από την παρούσα διατριβή χρησιμοποιεί στοχευόμενα την τρίτη ομάδα ως κατηγορία τεχνουργημάτων, τα εικονοστοιχεία της οποίας πρέπει να αποκλειστούν από την διαδικασία της ποσοτικοποίησης. Τέλος αξίζει να αναφερθεί η εργασία των Battiato *et al.* [153], η οποία είναι μια πολυκαναλική προσέγγιση θέτοντας τον αριθμό των ομάδων ίσο με 4, μια δηλαδή για κάθε χρώμα που εμφανίζεται σε μια έγχρωμη εικόνα μικροσυστοιχιών (κόκκινο, πράσινο, κίτρινο όταν έχουν υβριδοποιηθεί και τα δύο δείγματα με τον ανιχνευτή, και μαύρο όταν δεν έχει υβριδοποιηθεί κανένα από τα δύο δείγματα). Η προσέγγιση αυτή έχει περισσότερο ποιοτικό χαρακτήρα, εφόσον βοηθάει στην εύρεση των ανιχνευτών που υβριδοποιούνται ή δεν υβριδοποιούνται με τα δείγματα, ενώ δεν διαχωρίζει τα εικονοστοιχεία σήματος με τα εικονοστοιχεία υποβάθρου με σκοπό την εξαγωγή των ποσοτικοποιημένων τιμών.

3.3.2.2. Κατάτμηση με τον ασαφή αλγόριθμο C-μέσων

Στο ίδιο μήκος κύματος με την κατάτμηση με τον αλγόριθμο *K*-μέσων είναι βασισμένη και η κατάτμηση με τον ασαφή αλγόριθμό *C*-μέσων (Fuzzy *C*-means – FCM) [161]. Η μέθοδοι, οι οποίες κάνουν χρήση [155, 161] του ασαφούς αλγόριθμου *C*-μέσων δεν είναι πολλές στην βιβλιογραφία. Στην παρούσα διατριβή επιχειρήθηκε σύγκριση μεταξύ των δύο αλγορίθμων στο πεδίο της κατάτμηση μικροσυστοιχιών σύμφωνα με την οποία ο ασαφής αλγόριθμος *C*-μέσων επικρατεί του σαφούς αλγόριθμου *K*-μέσων. Η διαφορά των δύο μεθόδων απορρέει από το γεγονός ότι στην κατάτμηση με τον ασαφή αλγόριθμό *C*-μέσων το κάθε εικονοστοιχεία δεν αντιστοιχίζεται αυστηρά με κάποια από τις ομάδες, αλλά χαρακτηρίζεται από ένα ποσοστό συμμετοχής στην κάθε ομάδα. Τον αλγόριθμο θα αναπτύξουμε αναλυτικά στα επόμενα κεφάλαια.

3.3.2.3. Κατάτμηση με τυχαία Μαρκοβιανά πεδία

Μια άλλη μέθοδος [162-165] για τον διαχωρισμό των εικονοστοιχείων σήματος από τα εικονοστοιχεία υποβάθρου είναι η μέθοδος που στηρίζεται στα τυχαία Μαρκοβιανά πεδία (Markov Random Field - MRF). Η τεχνική αυτή κατάτμησης έχει χρησιμοποιηθεί σε διάφορα πεδία έρευνας, τα οποία απαιτούν επεξεργασία εικόνας, και είναι βασισμένη στην Μπεϋζιανή θεωρία. Αν υποθέσουμε ότι η εικόνα I είναι η πραγματοποίηση ενός τυχαίου πεδίου F , και ότι το c^* δηλώνει τον πραγματικό χαρακτηρισμό για ένα εικονοστοιχείο, τότε έστω ότι το \hat{c} θα είναι ένας προσδιορισμός του c^* . Η βασική ιδέα της μεθόδου είναι να προσδιοριστεί το c^* δοθέντος του y , το οποίο επιτυγχάνεται με την μεγιστοποίηση της εκ των υστέρων (*a posteriori*) πιθανότητας:

$$\hat{c} = \max_c p(I|c)p(c), \quad (3.5)$$

Η εκ των προτέρων πυκνότητα πιθανότητας (prior density) του c μοντελοποιείται από τον παρακάτω τύπο:

$$p(c) \propto \exp(\beta u), \quad (3.6)$$

Όπου β είναι μια θετική σταθερά, η οποία ελέγχει το μέγεθος των ομάδων των εικονοστοιχείων που θα δημιουργηθούν, ενώ το u συμβολίζει τον αριθμό των γειτόνων που ανήκουν στην ίδια ομάδα. Τέλος, η υπό συνθήκη πυκνότητα πιθανότητας της εικόνας μοντελοποιείται με μια Γκαουσιανή κατανομή:

$$p(c|I) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right) \exp\left(\frac{-(I-\mu)^2}{2\sigma^2} \right), \quad (3.7)$$

Όπου μ και σ είναι η μέση τιμή και η τυπική απόκλιση της Γκαουσιανής αντιστοίχως. Ο τύπος από τον οποίο μεγιστοποιείται η εκ των υστέρων πυκνότητα δίνεται ως εξής:

$$Q_i = \left(\frac{1}{2\sigma_i^2} \right) (I_0 - \mu_i)^2 - \beta u_i + \log(\sigma_i), \quad i = 1, \dots, k, \quad (3.8)$$

Όπου I_0 είναι η ένταση του εικονοστοιχείου, μ_i και σ_i είναι η μέση τιμή και η τυπική απόκλιση της κλάσης i και k είναι ο αριθμός των κλάσεων (στην περίπτωση των εικόνων

των μικροσυστοιχιών $k=2$). Εντοπίζοντας τα ελάχιστα της παραπάνω συνάρτησης τα εικονοστοιχεία χαρακτηρίζονται ως σήμα ή υπόβαθρο.

3.3.2.4. Κατάτμηση με διαμερισμό γύρω από το μέσο

Η κατάτμηση με διαμερισμό γύρω από το μέσο (Partition Around Medoids – PAM) [166] προτάθηκε ως μια πιο εξελιγμένη μέθοδος από την κατάτμηση με το αλγόριθμο K -μέσων. Σε αντίθεση με τον αλγόριθμο K -μέσων που ελαχιστοποιεί το τετραγωνικό σφάλμα των χαρακτηριστικών με το κεντροειδές, η κατάτμηση με διαμερισμό γύρω από το μέσο ελαχιστοποιεί ένα άθροισμα ανομοιοτήτας. Η μέθοδος πρώτα υπολογίζει τα μέσα (medoids), τα οποία χαρακτηρίζουν κάθε ομάδα, και τα οποία ελαχιστοποιούν την μέση ανομοιοότητα των εικονοστοιχείων που βρίσκονται στην ομάδα. Τέλος κάθε εικονοστοιχείο της εικόνας αντιστοιχίζεται στην ομάδα του μέσου που βρίσκεται πιο κοντά σε αυτό [154, 157-158].

3.3.2.5. Κατάτμηση με Μοντέλα Μεικτών Γκαουσιανών

Τέλος μια αρκετά διαδεδομένη μέθοδος [167-172] για την ομαδοποίηση των εικονοστοιχείων είναι η κατάτμηση με Μοντέλα Μεικτών Γκαουσιανών (Gaussian Mixture Model – GMM). Η μέθοδος αυτή θεωρεί ότι η κατανομή των φωτεινότητων μιας τυπικής κηλίδας στην εικόνα ακολουθεί ένα μείγμα κανονικών κατανομών. Οι ομάδες των εικονοστοιχείων επιλέγονται μεγιστοποιώντας την εκ των υστέρων πιθανότητα. Για τον σκοπό αυτό γίνεται χρήση του αλγόριθμου Προσδοκίας-Μεγιστοποίησης (EM) [173]. Η προσέγγιση αυτή μπορεί να αποδειχθεί ότι είναι περισσότερο αποδοτική από τους άλλους αλγόριθμους όταν οι ομάδες έχουν πολύ διαφορετικό μέγεθος, (γεγονός το οποίο μπορεί να συμβεί συχνά όταν χρησιμοποιούνται πάνω από δύο ομάδες), αλλά και όταν τα χαρακτηριστικά που χρησιμοποιούνται για κάθε εικονοστοιχείο δεν είναι ανεξάρτητα [170].

3.3.3.Υβριδική Μέθοδος

Πέραν των δύο μεγάλων κατηγοριών των μεθόδων για την κατάτμηση ο Rahnenführer και ο Bozinov [174-175] προσπάθησαν να συνδυάσουν τις δύο αυτές κατηγορίες υλοποιώντας μια υβριδική μέθοδο. Παρόμοια διαδικασία ακολούθησαν στο μέλλον και άλλες μέθοδοι [176]. Η υβριδική μέθοδος είναι η επέκταση μιας προηγούμενης εργασίας, η οποία ήταν βασισμένη στον αλγόριθμο ομαδοποίησης K-μέσων. Η συνεισφορά της μηχανικής μάθησης στην συγκεκριμένη μεθοδολογία έγκειται στην χρήση ομαδοποίησης για να προσεγγιστούν τα εικονοστοιχεία σήματος. Έχοντας ήδη αποφασίσει η μέθοδος έναν ικανό αριθμό εικονοστοιχείων σήματος, τοποθετώντας μια κυκλική μάσκα γύρο από την κάθε κηλίδα, τα εικονοστοιχεία που βρίσκονται εκτός αυτής ομαδοποιούνται σε δύο κατηγορίες, η περισσότερο φωτεινή από τις οποίες συμπεριλαμβάνονται τελικά στα εικονοστοιχεία σήματος.

3.4.Εξαγωγή Εντάσεων

Το τελικό στάδιο της επεξεργασίας εικόνων μικροσυστοιχιών περιλαμβάνει την εξαγωγή των εντάσεων της κάθε κηλίδας. Οι εντάσεις αυτές υπολογίζονται από τις φωτεινότητες των εικονοστοιχείων που έχουν χαρακτηριστεί ως σήμα ή υπόβαθρο. Σημειώνοντας τις εντάσεις σήματος με R_f και G_f αντίστοιχα για τις κόκκινες και τις πράσινες εντάσεις και R_b , G_b τις εντάσεις του υποβάθρου, λαμβάνουμε τις διορθωμένες εντάσεις ως:

$$R = R_f - R_b, \quad (3.9)$$

Και

$$G = G_f - G_b, \quad (3.10)$$

Ο λογαριθμικός λόγος των κόκκινων και πράσινων εντάσεων υπολογίζεται ως:

$$M = \log_2 \left(\frac{R}{G} \right), \quad (3.11)$$

Ενώ η λογαριθμική ένταση μιας κηλίδας, η οποία είναι μία ποσότητα που αντιπροσωπεύει την ολική ένταση της κηλίδας δίνεται ως εξής:

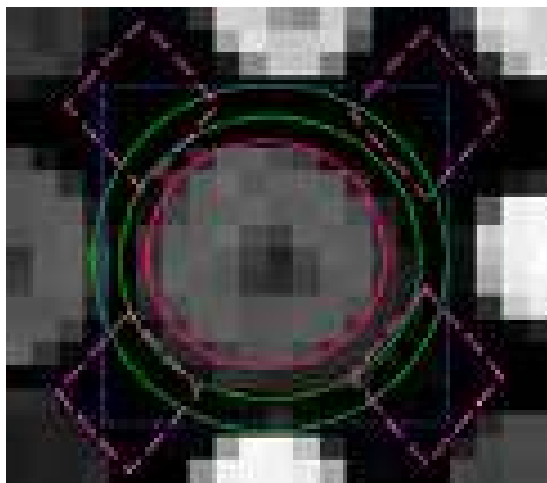
$$A = \frac{1}{2} \log_2 (R \cdot G), \quad (3.12)$$

Η επιλογή του λογάριθμου με βάση 2 γίνεται με σκοπό όταν διπλασιάζεται η ένταση να έχουμε μοναδιαία μεταβολή σε αυτά τα μεγέθη. Έτσι για παράδειγμα $M=0$ σημαίνει ότι έχουμε ίσες εντάσεις για τα δύο δείγματα, δηλαδή ότι τα δύο δείγματα έχουν εκφραστεί ισόποσα., $M=1$ σημαίνει διπλάσια διαφορά στις εντάσεις κ.ο.κ. Τέλος πολλές φορές χρησιμοποιείται ο απλός λόγος RAT μεταξύ των διορθωμένων τιμών των δύο καναλιών:

$$RAT = \frac{R}{G}. \quad (3.13)$$

Όπως και στην εξ. 3.5 θα συμβολίζεται με αστερίσκο η πραγματική τιμή μιας μεταβλητής (η τιμή από την επισημείωση), ενώ η τιμή που προσδιορίζεται από μια μέθοδο θα συμβολίζεται με μία «γωνία» από πάνω της. Έτσι τον λόγο RAT θα τον συναντήσουμε ως \hat{RAT} και RAT^* .

Τα πρώτα χρόνια που άρχισαν να αναπτύσσονται οι μέθοδοι επεξεργασίας των εικόνων των μικροσυστοιχιών συζήτηση γινόταν όσον αφορά στα εικονοστοιχεία που πρέπει να θεωρηθούν ως εικονοστοιχεία υποβάθρου. Το πρόγραμμα ScanAlyze [75] θεωρεί ως εικονοστοιχεία υποβάθρου όλα τα εικονοστοιχεία που βρίσκονται εκτός του σταθερού κύκλου. Μία άλλη προσέγγιση χρησιμοποίησε το πρόγραμμα Quantaray [136], σύμφωνα με το οποίο θεωρείται ένας δακτύλιος ομόκεντρος με την κηλίδα μέσα από τον οποίο λαμβάνονται τα εικονοστοιχεία υποβάθρου. Η προσέγγιση αυτή ήταν περισσότερο αποδοτική από την προηγούμενη, διότι περιορίζει τον κίνδυνο να συμπεριληφθούν στα εικονοστοιχεία υποβάθρου, τα εικονοστοιχεία που είναι πολύ κοντά στα εικονοστοιχεία σήματος (των εικονοστοιχείων αυτών η φωτεινότητα είναι επηρεασμένη από την υψηλή φωτεινότητα των εικονοστοιχείων σήματος). Τέλος υπήρχε και η προσέγγιση, την οποία χρησιμοποιεί το πρόγραμμα Genepix [130] λαμβάνοντας τα εικονοστοιχεία υποβάθρου από τις τετραγωνικές περιοχές ανάμεσα στις γειτονικές κηλίδες. Αυτές οι περιοχές είναι οι πιο απομακρυσμένες περιοχές από κάθε κηλίδα. Οι τρεις προσεγγίσεις αυτές αποτυπώνονται στο σχήμα 3.13.



Σχήμα 3.13: Οι περιοχές από τις οποίες λαμβάνονταν τα εικονοστοιχεία υποβάθρου από τις πρώτες μεθόδους κατάτμησης.

Με την πάροδο του χρόνου και ενώ οι πιο εξελιγμένες μέθοδοι κατάτμησης έκαναν την εμφάνιση τους, οι προσεγγίσεις αυτές έπαψαν να χρησιμοποιούνται, καθώς τα εικονοστοιχεία σήματος και τα εικονοστοιχεία υποβάθρου προσδιορίζονται σαφώς. Οι εντάσεις σήματος της κάθε κηλίδας R_f και G_f συνήθως υπολογίζονται ως η μέση φωτεινότητα των εικονοστοιχείων σήματος, ενώ η διάμεσος τιμή των εντάσεων των εικονοστοιχείων υποβάθρου ορίζει την ένταση του υποβάθρου της κάθε κηλίδας.

Η συνεισφορά της παρούσας διατριβής για εξαγωγή της έντασης των κηλίδων απορρέει έμμεσα, όπως θα δούμε σε επόμενα κεφάλαια, από το γεγονός ότι όλες οι μέθοδοι κατάτμησης υλοποιήθηκαν με γνώμονα την εξαίρεση τεχνουργημάτων αλλά και άλλων φωτεινών εικονοστοιχείων, από τα εικονοστοιχεία σήματος και υποβάθρου. Έτσι τα εικονοστοιχεία από τα οποία υπολογίζονται τα παραπάνω μεγέθη και εξάγεται η ποσοτικοποίηση της υβριδοποίησης, είναι τα αντιπροσωπευτικότερα από τις ομάδες του σήματος και του υποβάθρου.

3.5.Συμπεράσματα

Στο κεφάλαιο αυτό έγινε μια αναλυτική επισκόπηση της βιβλιογραφίας στο πεδίο της επεξεργασίας των εικόνων μικροσυστοιχιών. Σε διάφορα σημεία του κεφαλαίου έγινε αντιληπτό ότι η παρούσα επιστημονική στάθμιση δίνει μεγάλα περιθώρια βελτίωσης των

μεθόδων. Η ανάγκη γενικευμένων αλλά και ευφυών προσεγγίσεων στον εντοπισμό των κηλίδων, με σκοπό να αντιμετωπιστούν η διαφορετικότητα των μικροσυστοιχιών αλλά και ο συνεχώς αυξανόμενος αριθμός των ανιχνευτών, είναι ένα πεδίο έρευνας. Για παράδειγμα, φαίνεται ελκυστική μια γενικευμένη μέθοδος, η οποία να μπορεί να επεξεργαστεί εικόνες μικροσυστοιχιών, οι κηλίδες των οποίων χωροθετούνται είτε σε τετραγωνικό είτε σε εξαγωνικό πλέγμα, και η οποία αναπτύχθηκε στην παρούσα διατριβή. Ένα άλλο αντικείμενο που επιδέχεται μεγάλη βελτίωση είναι η σωστότερη επιλογή των εικονοστοιχείων σήματος και υποβάθρου κατά την διαδικασία της κατάτμησης. Έχει μεγάλη σημασία για την εξαγωγή της έντασης μιας κηλίδας, να γίνεται λαμβάνοντας υπόψη στους υπολογισμούς τα σωστά εικονοστοιχεία. Λανθασμένος υπολογισμός της έντασης μια κηλίδας οδηγεί άμεσα σε λανθασμένα συμπεράσματα για την βιολογική συμπεριφορά του ανιχνευτή. Τα παραπάνω σημεία είναι και αυτά τα οποία απασχόλησαν την παρούσα διδακτορική διατριβή. Όπως θα δούμε στα επόμενα κεφάλαια, η παρούσα διατριβή συνεισέφερε καινοτόμες λύσεις στα παραπάνω ζητήματα.

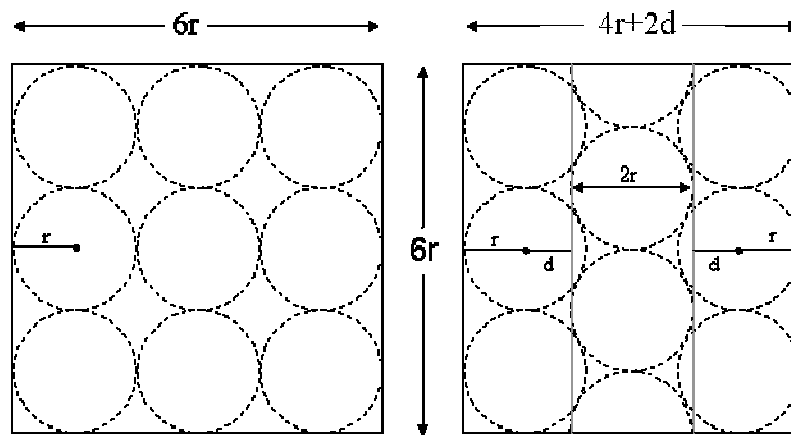
Κεφάλαιο 4ο: Εντοπισμός Κηλίδων και τοποθέτηση πλέγματος

4.Εντοπισμός Κηλίδων και τοποθέτηση πλέγματος

4.1.Εισαγωγή

Η διαδικασία της εύρεσης των κηλίδων αποτελεί την έναρξη της επεξεργασίας των εικόνων των μικροσυστοιχιών. Είναι και αυτή μια διαδικασία η οποία είναι εξαιρετικά κρίσιμη, διότι οποιαδήποτε αποτυχία της μεταφέρεται αυτόματα και στα επόμενα στάδια της επεξεργασίας των εικόνων. Στο κεφάλαιο αυτό περιγράφονται όλες οι τεχνικές που χρησιμοποιήθηκαν για το στάδιο αυτό, με σκοπό να καταλήξουμε σε μια ενιαία μεθοδολογία. Η γενίκευση των μεθόδων έγινε για την αντιμετώπιση της διαφορετικότητας στην φύση των εικόνων όσον αφορά στο πλέγμα το οποίο είναι τοποθετημένα τα βιολογικά αντικείμενα επάνω στην γυάλινη επιφάνεια (και κατ' επέκταση οι κηλίδες στην παραγόμενη εικόνα). Υπάρχουν δηλαδή δύο διαφορετικά ήδη μικροσυστοιχιών, το ένα έχει τοποθετημένες τις κηλίδες σε γραμμές και στήλες ενώ στο άλλο οι κηλίδες είναι τοποθετημένες με εξαγωνική δομή [114-115]. Δηλαδή οι δύο διαφορετικοί τύποι

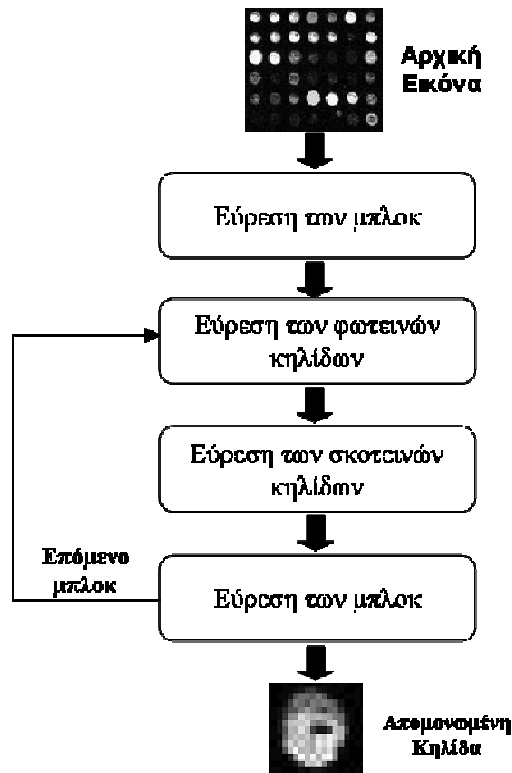
μικροσυστοιχιών χρησιμοποιούν είτε τετραγωνικό πλέγμα είτε εξαγωνικό πλέγμα αντίστοιχα. Η πρώτες μικροσυστοιχίες δημιουργήθηκαν με τετραγωνικό πλέγμα, όσο όμως η τεχνολογία εξελισσόταν και όσο η απαίτηση για εξέταση μεγαλύτερου αριθμού βιολογικών δεδομένων μεγάλωνε, δημιουργήθηκε η ανάγκη για περισσότερους ανιχνευτές σε μία μικροσυστοιχία. Έτσι, για την εξοικονόμηση χώρου εφαρμόστηκε το εξαγωνικό πλέγμα για την χωροθέτηση των ανιχνευτών, όπως φαίνεται και στο σχήμα 4.1. Μάλιστα μικραίνοντας και λίγο την ακτίνα της κάθε κηλίδας ο χώρος που μπορεί να εξοικονομηθεί αυξάνεται σημαντικά.



Σχήμα 4.1: Σύγκριση μεταξύ του τετραγωνικού και του εξαγωνικού πλέγματος

Ο πρώτος στόχος της παρούσας διατριβής ήταν η δημιουργία ενός αλγορίθμου, ο οποίος να επεξεργάζεται εικόνες με εξαγωνικό πλέγμα, δεδομένου ότι μέχρις στιγμής στην βιβλιογραφία είχε παρουσιαστεί μόνο μια εργασία. Επίσης η χωροθέτηση των ανιχνευτών σε εξαγωνικό πλέγμα θα χρησιμοποιείται ολοένα και περισσότερο για την εξοικονόμηση χώρου στην μικροσυστοιχία. Στην συνέχεια το ζητούμενο ήταν η γενίκευση της προτεινόμενης μεθοδολογίας ώστε να μπορεί να λειτουργήσει και στους δύο τύπους πλέγματος.

Γενικά οι μεθοδολογία ακολουθεί τυποποιημένα βήματα σαν αυτά που είδαμε στο προηγούμενο κεφάλαιο. Το σχήμα 4.2 παρουσιάζει σε πρώτο επίπεδο το διάγραμμα ροής της μεθοδολογίας.

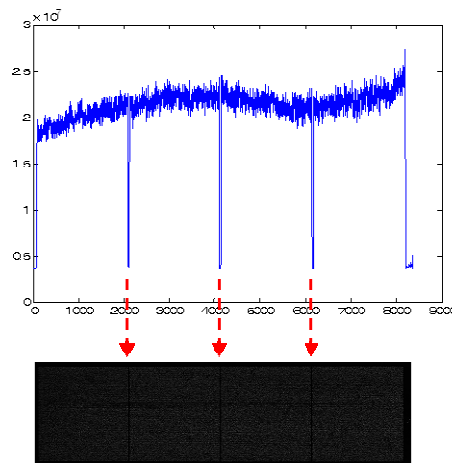


Σχήμα 4.2: Τα τέσσερα βήματα σταδίου της εύρεσης των κηλίδων και της τοποθέτηση του πλέγματος

Ένα βασικό πρώτο βήμα είναι ο διαχωρισμός των συστάδων της εικόνας. Η εικόνα θα διαχωριστεί στις επιμέρους συστάδες, οι οποίες στην συνέχεια θα επεξεργαστούν η καθεμία ξεχωριστά. Στην συνέχεια εντοπίζονται όλα τα φωτεινά αντικείμενα της εικόνας, τα οποία και πιθανότατα είναι κηλίδες. Πέραν όμως των φωτεινών κηλίδων που υπάρχουν στην εικόνα και που εντοπίζονται σχετικά εύκολα λόγω της υψηλής τους φωτεινότητας, μια αυτόματη μέθοδος πρέπει να εντοπίζει και τις θέσεις των κηλίδων εκείνων που έχουν πολύ μικρή φωτεινότητα και οι οποίες προέρχονται από την ελλιπή υβριδοποίηση των δειγμάτων με τους ανιχνευτές κατά την διάρκεια του βιολογικού πειράματος. Τέλος, εφόσον η μέθοδος έχει εντοπίσει τις θέσεις όλων των κηλίδων στην εικόνα ένα πλέγμα τοποθετείται με σκοπό να απομονώσει την κάθε κηλίδα μέσα σε ένα κελί. Τα κελιά αυτά που περιέχουν μία μόνο κηλίδα είναι αυτά που θα υποστούν την κατάτμηση κάθε ένα ξεχωριστά, ώστε τελικά να βρεθούν τα εικονοστοιχεία σήματος και τα εικονοστοιχεία υποβάθρου για την κάθε κηλίδα.

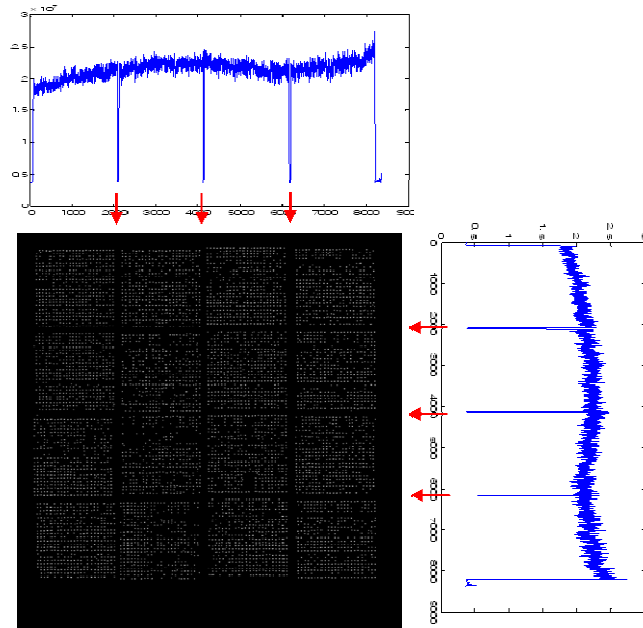
4.2.Εύρεση των Συστάδων

Η εύρεση των συστάδων, η οποία υλοποιήθηκε κατά την διάρκεια της διατριβής είναι βασισμένη στην ολιστική προσέγγιση όπως αυτή αναφέρθηκε στο 3ο Κεφάλαιο. Ανεξάρτητα από το πλέγμα το οποίο χρησιμοποιήθηκε για την κατασκευή της μικροσυστοιχίας, οι προβολές της εικόνας μπορούν εύκολα να αναδείξουν τις συστάδες της εικόνας. Από τις προβολές της εικόνας θα βρεθούν τα σημεία στα οποία θα χωριστεί η εικόνα στις επιμέρους συστάδες. Μια εικόνα μικροσυστοιχιών ωστόσο μπορεί να έχει γραμμές και στήλες από συστάδες, ή μόνο μια γραμμή από συστάδες, ή μόνο μια στήλη από συστάδες. Αντίστοιχα, στις τρεις αυτές περιπτώσεις είτε είναι απαραίτητες και η οριζόντια και η κατακόρυφη προβολή της εικόνας, είτε μόνο η κατακόρυφη, είτε μόνο η οριζόντια.



Σχήμα 4.3: Κατακόρυφη προβολή μια εικόνας μικροσυστοιχιών η οποία περιέχει τέσσερις συστάδες σε μια γραμμή, η μια δίπλα στην άλλη.

Παραδείγματος χάριν στο σχήμα 4.3 παρουσιάζεται μια εικόνα μικροσυστοιχιών, η οποία έχει συστάδες μόνο σε οριζόντια παράταξη (σε μια γραμμή δηλαδή από συστάδες), με αποτέλεσμα για την εύρεση των κοιλάδων που σημαίνονται με τα βέλη να απαιτείται μόνο η κατακόρυφη προβολή της εικόνας. Αντιθέτως, το παρακάτω σχήμα 4.4 δείχνει μια εικόνα, η οποία περιέχει και γραμμές και στήλες από συστάδες, και κατά συνέπεια είναι απαραίτητες και οι δύο προβολές.



Σχήμα 4.4: Εικόνα μικροσυστοιχιών η οποία περιέχει γραμμές και στήλες από συστάδες.

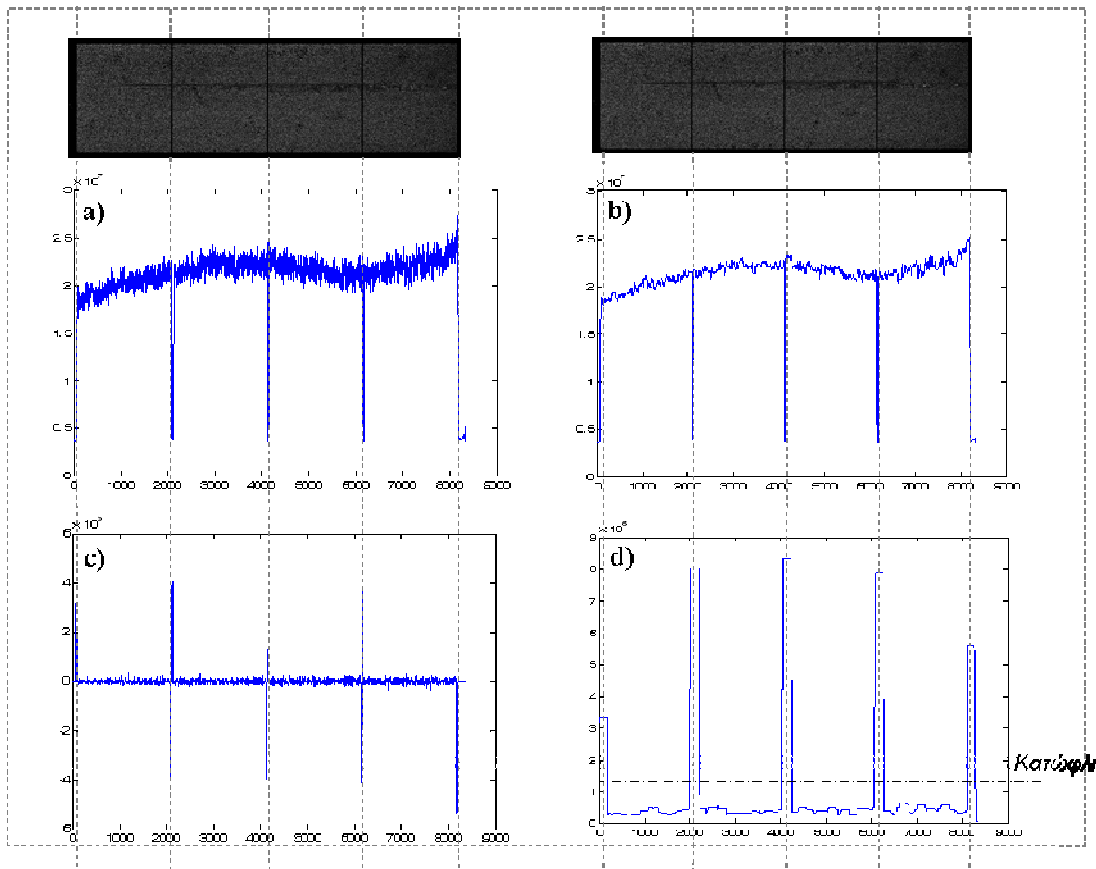
4.2.1. Αρχική Προσέγγιση

Σε πρώτο στάδιο, θέλοντας να δημιουργήσουμε μια μέθοδο για τον εντοπισμό των κηλίδων σε εικόνες τετραγωνικού πλέγματος, επεξεργαστήκαμε τις δύο προβολές τις εικόνας, με σκοπό να γίνουν πολύ πιο ευδιάκριτες από ότι ήταν αρχικά. Αρχικά οι προβολές εξομαλύνθηκαν με χρήση ενός φίλτρου διαμέσου. Όπως φαίνεται στο σχήμα 4.5^α, μετά την εφαρμογή του φίλτρου η κορυφές στο σήμα εξομαλύνονται αρκετά, ενώ οι κοιλάδες μένουν ανέπαφες. Αντί να εντοπίσουμε τις κοιλάδες σε αυτήν την φάση, είναι περισσότερο βολικό να εντοπιστούν οι κορυφές στην κλίση (gradient) του σήματος. Κατ' αυτόν τον τρόπο, η μέθοδος πλέον θα εντοπίζει τα σημεία στο σήμα, στα οποία η κλίση του σήματος αυξομειώνεται απότομα.

Γενικά η κλίση οποιουδήποτε μονοδιάστατου σήματος δίνεται από την σχέση:

$$\nabla Proj = \frac{dProj}{dx} i, \quad (4.1)$$

Όπου *Proj* είναι μια από τις προβολές της εικόνας. Η κλίση του σήματος απεικονίζεται στο σχήμα 4.5.γ.



Σχήμα 4.5: Τα αποτελέσματα των τεσσάρων βημάτων της επεξεργασίας των προβολών της εικόνας.

Για να εντοπιστούν οι κορυφές στην κλίση του σήματος, οι οποίες αντιστοιχούν στα σημεία που θα χωριστεί οι εικόνα στις επιμέρους συστάδες, το μονοδιάστατο σήμα της κλίση της προβολής σαρώνεται παραθυρικά με ένα παράθυρο 25 σημείων. Ένα καινούργιο σήμα δημιουργείται όπως φαίνεται στο σχήμα 4.5δ. Το εκάστοτε κεντρικό σημείο του συρόμενου παραθύρου απόκτα τιμή ίση με την διαφορά του μέγιστου μείων το ελάχιστο που περιλαμβάνονται σε αυτό το παράθυρο. Η τιμή του σήματος στο σημείο i δηλαδή δίνεται από την σχέση:

$$W_{ID}(i) = \max_{window}(i) - \min_{window}(i), \quad (4.2)$$

Όπως παρατηρούμε στο σχήμα 4.5.δ οι κορυφές του σήματος έχουν γίνει πλέον ευδιάκριτες και εύκολα διαχωρίσιμες από το υπόλοιπο σήμα με μια απλή κατοφλίωση. Είναι σημαντικό να αναφερθεί ότι το μέγεθος του παραθύρου μπορεί να φαίνεται πολύ

αυστηρά καθορισμένο, είναι αποδεκτό όμως δεδομένου ότι τα μεγέθη των κηλίδων είναι της ίδιας τάξης μεγέθους για οποιαδήποτε εικόνα μικροσυστοιχιών, ενώ το ίδιο ισχύει και για τις αποστάσεις μεταξύ των συστάδων. Επίσης μικρές αποκλίσεις στο μέγεθος του παραθύρου αυτού δεν αλλάζει τίποτα δραματικά ως προς τον διαχωρισμό των συστάδων.

Στην συνέχεια όλα τα υπόλοιπα σημεία του τελικού σήματος απαλείφονται με την εφαρμογή ενός κατώφλιού. Δεδομένου ότι το μέγιστο έχουν γίνει εξέχοντα εφαρμόζοντας όλα τα παραπάνω βήματα ένα απλό κατώφλι είναι αρκετό για να διαχωρίσουμε της κορυφές του σήματος από τα υπόλοιπα σημεία. Στην περίπτωση μας το κατώφλι ορίστηκε να είναι:

$$Th_l = 2 * mean(W_{ID}), \quad (4.3)$$

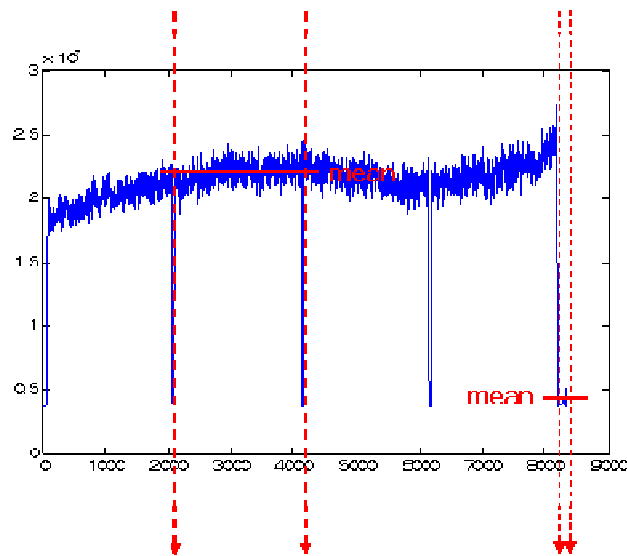
Ο εντοπισμός των συστάδων τελειώνει με την επαλήθευση των σημείων διαχωρισμού που έχουν βρεθεί. Όπως φαίνεται και στο σχήμα 4.5δ εκτός των σημείων που έχουν βρεθεί και τα οποία αντιστοιχούν στις κοιλάδες, έχουν εντοπιστεί και κορυφές κοντά στα όρια της εικόνας τα οποία αντιστοιχούν στα σημεία που ξεκινάει η πρώτη συστάδα και που τελειώνει η τελευταία συστάδα. Οι κορυφές αυτές είναι μικρότερες από τις άλλες λόγω του γεγονός ότι δεν είναι σημεία τα οποία βρίσκονται ανάμεσα σε δύο συστάδες αλλά είναι δίπλα μόνο σε μία. Αυτό το γεγονός μπορεί να δημιουργήσει πρόβλημα στην μέθοδο γιατί καμιά φορά οι κορυφές αυτές είναι τόσο μικρές ώστε να απαλείφονται από το κατώφλι Th_l . Για τον λόγο αυτό στα τελικά σημεία προς επαλήθευση προσθέτουμε και τα άκρα της εικόνας. Τα σημεία αυτά εξετάζονται ανά δύο, ώστε τα σημεία τα οποία έχουν μέση τιμή σήματος ανάμεσα τους μεγαλύτερη από το κατώφλι Th_l είναι και αυτά τα οποία τελικά θα αποτελέσουν τα σημεία διαχωρισμού. Η διαδικασία επαλήθευσης των σημείων διαχωρισμού φαίνεται στο σχήμα 4.6. Όπως φαίνεται στο σχήμα η μέση τιμή του αρχικού σήματος ανάμεσα από το ζευγάρι που είναι κοντά στις τιμές (~2000,~4000) είναι υψηλή. Αντίθετα, τα δύο σημεία κοντά στις τιμές 8000, έχουν ανάμεσα τους μέση τιμή σήματος πολύ χαμηλή και για τον λόγο αυτό απορρίπτονται.

Στην περίπτωση που υπάρχουν στην εικόνα σειρές και στήλες από συστάδες, δημιουργούνται δύο διανύσματα V και H για την κατακόρυφη και για την οριζόντια προβολή αντίστοιχα:

$$V = [1, v_1, v_2, \dots, v_{N_v}, D_v], \quad (4.4)$$

$$H = [1, h_1, h_2, \dots, h_{N_h}, D_h], \quad (4.5)$$

όπου D_v και D_h είναι οι διαστάσεις της εικόνας στην κατακόρυφο και την οριζόντια κατεύθυνση, αντίστοιχα, ενώ N_v και N_h είναι οι αριθμοί των κοιλάδων που έχουν εντοπιστεί στις δύο κατευθύνσεις. Τα δύο αυτά διανύσματα περιέχουν τις συντεταγμένες στις οποίες θα χωριστεί η εικόνας με σκοπό να απομονωθούν οι επιμέρους συστάδες.

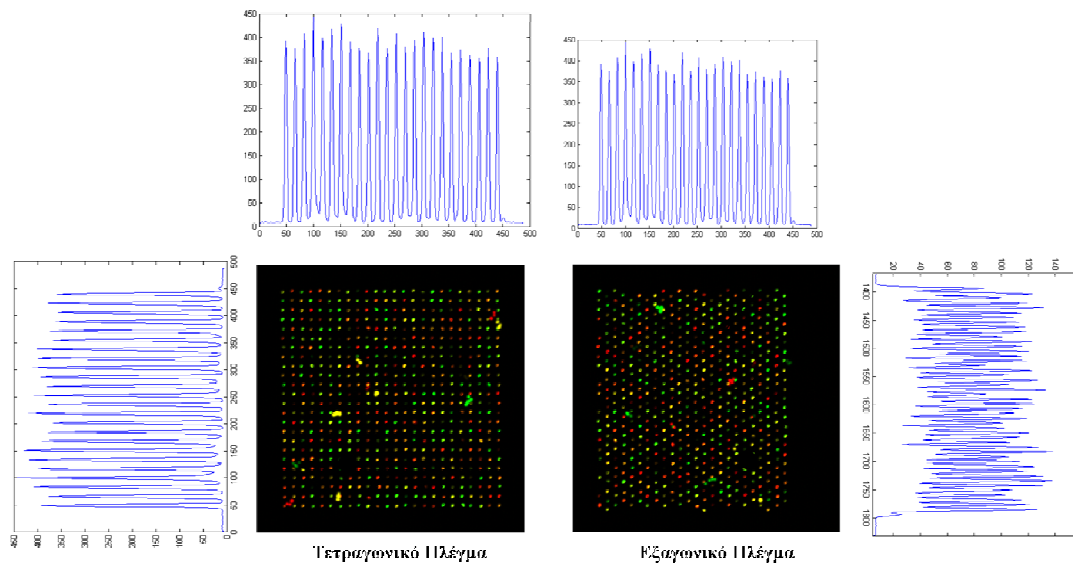


Σχήμα 4.6: Η διαδικασία της επαλήθευσης των κοιλάδων που έχουν εντοπιστεί.

4.2.2.Γενικευμένη προσέγγιση

Επιχειρώντας την γενίκευση της μεθόδου για να επεξεργάζεται και τους δύο τύπους εικόνων, διαπιστώσαμε ότι οι εικόνες που κάνουν χρήση εξαγωνικού πλέγματος παράγουν προβολές με διαφορετικά χαρακτηριστικά. Στο σχήμα 4.7 φαίνεται αυτή η διαφορά. Όπως παρατηρείται οι δύο προβολές (οριζόντια και κατακόρυφη) μιας εικόνας με τετραγωνικό

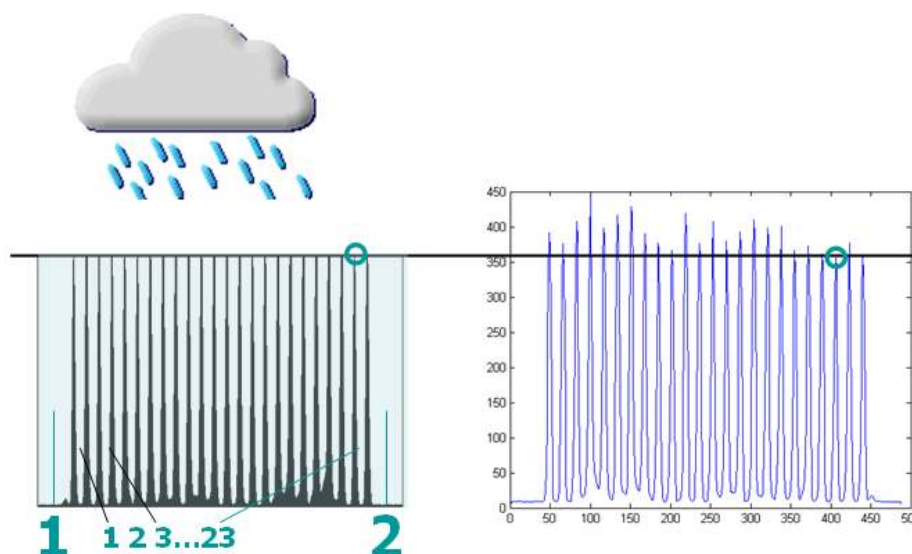
πλέγμα είναι παρόμοιες. Σε αυτές υπάρχουν υψηλές κορυφές και «βαθιές» κοιλάδες, οι οποίες φτάνουν περίπου μέχρι το μηδέν. Σε αναντιστοιχία ενώ η κατακόρυφη προβολή της εικόνας με εξαγωνικό πλέγμα έχει και αυτή παρόμοια χαρακτηριστικά με τις προαναφερθείσες, η οριζόντια προβολή είναι διαφορετική. Η κοιλάδες σε αυτήν την περίπτωση δεν είναι το ίδιο «βαθιές». Αυτό προκύπτει από το γεγονός ότι η κηλίδες της εικόνας τώρα δεν είναι χωροθετημένες σε γραμμές (η μια δίπλα στην άλλη).



Σχήμα 4.7: Οι προβολές μια εικόνας τετραγωνικού πλέγματος και μια εικόνας εξαγωνικού πλέγματος.

Η παραπάνω διαφοροποίηση στην φύση των προβολών του εξαγωνικού πλέγματος δεν αφήνει περιθώρια στην προηγούμενη προσέγγιση να λειτουργήσει σωστά σε εικόνες με εξαγωνικό πλέγμα. Για να γενικευθεί λοιπόν όλη η μεθοδολογία του εντοπισμού των κηλίδων και της τοποθέτησης του πλέγματος απαιτήθηκε η αντικατάσταση του βήματος αυτού με μια διαδικασία που να λειτουργεί και στους δύο τύπους εικόνων. Το βασικό πρόβλημα το δημιουργεί το κατώφλι Th_1 , η επιλογή του οποίου πλέον περιπλέκεται πολύ. Για τον λόγο αυτό υλοποιήθηκε ένας αλγόριθμος ο οποίος είναι εμπνευσμένος από τον αλγόριθμο του υδροκριτή [142], με σκοπό να αποφευχθεί η χρήση κατωφλίωσης. Θεωρήστε λοιπόν την προβολή της εικόνας σαν την κατακόρυφη τομή μια τοπογραφικής περιοχής, και ότι στην περιοχή αυτή ξεκινάει να βρέχει. Οι κοιλάδες της προβολής αρχίζουν να γεμίζουν με νερό. Η επιφάνεια του νερού που συνέχεια ανεβαίνει λειτουργεί ως ένα κινητό κατώφλι το οποίο θα σταματήσει να ανεβαίνει μόλις μια από τις κοιλάδες ξεχειλίσει σε μια από τις δύο διπλανές τις. Η διαδικασία αυτή φαίνεται στο σχήμα 4.8.

Όταν η προβολή γεμίσει με νερό, από την ποσότητα του νερού προσδιορίζεται ο χωρητικότητα της κάθε μιας κοιλάδας. Όπως αναφέρθηκε και προηγουμένως υπάρχουν δύο τύποι κοιλάδων. Οι μεγάλες κοιλάδες οι οποίες σχηματίζονται από τις περιοχές ανάμεσα στις συστάδες της εικόνας, και οι μικρές κοιλάδες οι οποίες δημιουργούνται από τις περιοχές ανάμεσα στις κηλίδες. Για τον διαχωρισμό των συστάδων πρέπει να εντοπιστούν οι μεγάλες κοιλάδες και να διαχωριστούν από τις μικρές.



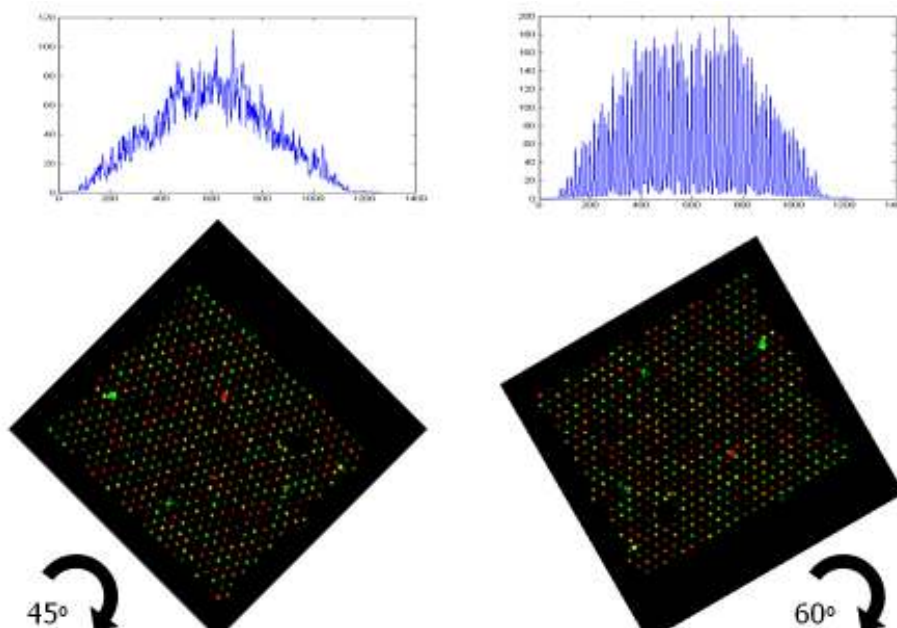
Σχήμα 4.8: Καθώς το νερό γεμίζει την προβολή της εικόνας η επιφάνεια του λειτουργεί ως μεταβλητό κατώφλι. Τελικά το νερό σταματάει όταν μια κοιλάδα ξεχειλίζει από νερό (στο σημείο που επισημαίνεται με τον κύκλο. Στην προβολή δημιουργούνται μικρές και μεγάλες κοιλάδες.

Για το σκοπό αυτό η χωρητικότητα της κάθε κοιλάδας χρησιμοποιείται ως χαρακτηριστικό και τροφοδοτούν τον αλγόριθμο K-μέσων [152] ο οποίος διαχωρίζει τις συστάδες σε δύο ομάδες. Από τις συντεταγμένες των μεγάλων κοιλάδων στο σήμα των προβολών εξάγονται οι συντεταγμένες των σημείων που διαχωρίζουν την εικόνα στις επιμέρους συστάδες.

Η μεθοδολογία εκμεταλλεύεται την ομαδοποίηση των κοιλάδων και για τον προσδιορισμό της απόστασης δύο γειτονικών κηλίδων. Η απόσταση αυτή θα χρειαστεί για το βήμα του εντοπισμού των μη υβριδοποιημένων κηλίδων. Για τον υπολογισμό της γίνεται χρήση των κοιλάδων με μικρή χωρητικότητα αφού αυτές είναι που προκύπτουν από τις περιοχές στην εικόνα που διαχωρίζουν τις κηλίδες. Έτσι για κάθε ζεύγος διαδοχικών κοιλάδων χαμηλής χωρητικότητας υπολογίζεται η απόσταση τους στο σήμα των προβολών. Την επικρατούσα

τιμή όλων αυτών των αποστάσεων που υπολογίζονται, την ονομάζουμε επικρατούσα απόσταση (*mode_dist*) και είναι αυτή η οποία θα χρησιμοποιήσουμε αργότερα για να προσομοιώσουμε την απόσταση μεταξύ δύο γειτονικών κηλίδων.

Στην γενικευμένη προσέγγιση είναι απαραίτητο για την μετέπειτα διαδικασία να γίνει η αναγνώριση του πλέγματος που χρησιμοποιεί η εικόνα. Ένας έξυπνος τρόπος για να επιτευχθεί ο στόχος αυτός είναι η διερεύνηση της προβολής της εικόνας στραμμένης κατά 45° και κατά 60° . Εάν μια εικόνα τετραγωνικού πλέγματος στραφεί κατά 45° οι κηλίδες της θα βρεθούν διατεταγμένες σε κατακόρυφη διάταξη. Αντίστοιχα σε κατακόρυφη διάταξη θα βρεθούν οι κηλίδες σε μια εικόνα εξαγωνικού πλέγματος εάν αυτή στραφεί στις 60° , όπως φαίνεται στο σχήμα 4.9 στα δεξιά. Στρέφοντας λοιπόν την προς εξέταση εικόνα στις παραπάνω μοίρες και εξάγοντας τις προβολές της σε αυτήν την διεύθυνση παρατηρούμε ότι η φύση των δύο προβολών είναι τελείως διαφορετική.



Σχήμα 4.9: Μια εικόνα εξαγωνικού πλέγματος στραμμένη στις 45° και στις 60° , και οι προβολές της σε αυτές τις διευθύνσεις..

Όπως φαίνεται στο σχήμα 4.9 η τυπική απόκλιση της προβολής στις 60° μιας εικόνας εξαγωνικού πλέγματος είναι πολύ μεγαλύτερη από αυτή στις 45° , δεδομένου ότι το σήμα παρουσιάζει μεγάλες διακυμάνσεις. Σύμφωνα με το γεγονός αυτό η μέθοδος μπορεί

εύκολα να αποφανθεί αυτόματα ποιο είναι το πλέγμα που χρησιμοποιεί η κάθε προς επεξεργασία εικόνα από τον παρακάτω κανόνα:

$$\begin{aligned} \text{IF std}(60^\circ) > \text{std}(45^\circ) \text{ THEN "hex"} \\ \text{IF std}(45^\circ) > \text{std}(60^\circ) \text{ THEN "rec"} \end{aligned} \quad (4.6)$$

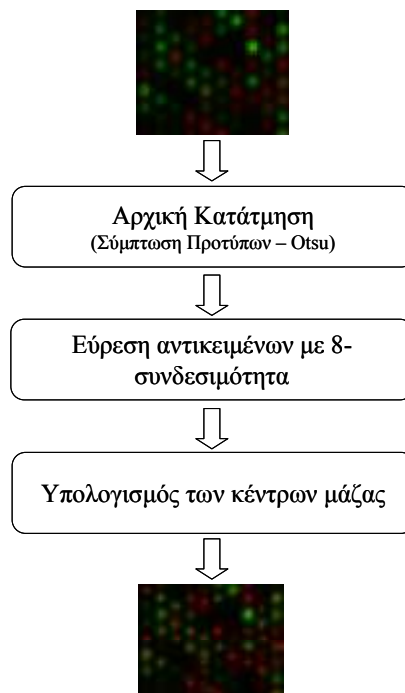
Η παραπάνω απόφαση θα χρησιμοποιηθεί στο βήμα του εντοπισμού των μη υβριδοποιημένων κηλίδων.

4.3. Εντοπισμός των υβριδοποιημένων κηλίδων

Μετά τον διαχωρισμό της εικόνας στις επιμέρους συστάδες ακολουθεί ο εντοπισμός των φωτεινών αντικειμένων σε καθεμιά από τις συστάδες. Τα φωτεινά αυτά αντικείμενα είναι πιθανότατα κηλίδες, οι οποίες αντιστοιχούν σε ανιχνευτές που έχουν υβριδοποιηθεί με τα δείγματα. Εκτός όμως από φωτεινές κηλίδες, αντικείμενα μπορεί να εντοπιστούν και από διάφορα τεχνουργήματα που υπάρχουν στην εικόνα. Τα αντικείμενα αυτά μάλιστα μπορεί να είναι παρόμοιου μεγέθους με τις κηλίδες. Ο εντοπισμός των φωτεινών αντικειμένων είναι κατά κανόνα βασισμένος σε μεθόδους κατάτμησης της εικόνας [177], με την διαφορά ότι λαμβάνει χώρα μόνο για τον εντοπισμό των αντικειμένων και η ακρίβεια του σε επίπεδο εικονοστοιχείων δεν είναι ο κυριότερος στόχος. Άλλωστε κάθε κηλίδα θα κατατμηθεί με ακρίβεια στο επόμενο στάδιο της επεξεργασίας των εικόνων των μικροσυστοιχιών, όπως θα δούμε στο επόμενο κεφάλαιο. Επίσης από το συγκεκριμένο βήμα δεν προσδοκούμε την εύρεση όλων των κηλίδων της εικόνας αφού εξ αρχής γνωρίζουμε ότι στην εικόνα θα υπάρχουν και σκοτεινές κηλίδες από ανιχνευτές, οι οποίοι δεν έχουν υβριδοποιηθεί με τα δείγματα. Ωστόσο, όσο μεγαλύτερος είναι ο αριθμός των σωστά εντοπισμένων κηλίδων από το βήμα αυτό, τόσο πιο εύκολος θα είναι ο εντοπισμός των μη υβριδοποιημένων κηλίδων. Η μέθοδος που προτείνεται για αυτήν την διαδικασία ακολουθεί τρία βήματα όπως φαίνεται και στο σχήμα 4.10.

Μια μέθοδος κατάτμησης εφαρμόζεται αρχικά με σκοπό να μετατρέψουμε την εικόνα σε δυαδική και να ξεχωρίσουμε τα φωτεινά αντικείμενα μέσα στην εικόνα. Η κατάτμηση αυτή εφαρμόζεται σε ολόκληρη την εικόνα και δεν έχει καμιά σχέση με την κατάτμηση

που εφαρμόζεται αργότερα σε κάθε μια κηλίδα ξεχωριστά για τον διαχωρισμό των εικονοστοιχείων σήματος από τα εικονοστοιχεία υποβάθρου. Είναι σημαντικό επειδή μια εικόνα μικροσυστοιχιών είναι μεγάλη σε μέγεθος η τεχνική για την αρχική αυτή κατάτμηση που θα εφαρμοστεί να είναι αποτελεσματική, γρήγορη και να απαιτεί μικρή υπολογιστική ισχύ. Δύο τεχνικές είναι αυτές οι οποίες εφαρμόστηκαν για αυτό τον σκοπό. Αρχικά χρησιμοποιήσαμε την τεχνική σύμπτωσης προτύπου (Template Matching) [113], ενώ σε μια άλλη εργασία χρησιμοποιήθηκε η τεχνική κατάτμησης του Ότσου (Otsu) [112].

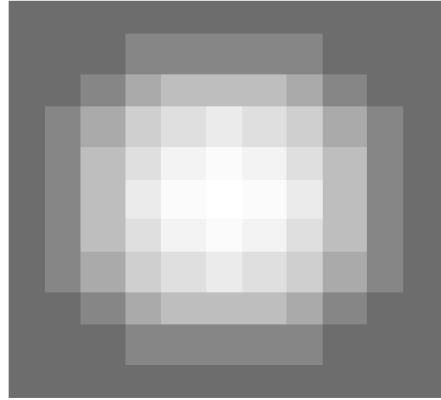


Σχήμα 4.10: Διάγραμμα ροής για τον εντοπισμό των υβριδοποιημένων κηλίδων.

Η σύμπτωσης προτύπου είναι μία τεχνική, κατά την οποία σαρώνεται ολόκληρη η εικόνα, εντοπίζοντας αντικείμενα που έχουν μεγάλη ομοιότητα με το πρότυπο που χρησιμοποιείται. Η επιλογή του προτύπου στην προκειμένη περίπτωση αποτελεί σίγουρα ένα θέμα περαιτέρω διερεύνησης. Το καταλληλότερο πρότυπο για να προσομοιώσει μια κηλίδα [92,114-115], όπως θεωρητικά αυτή θα έπρεπε να είναι, είναι μια Γκαουσιανή κατανομή δύο διαστάσεων με μέγεθος περίπου ίσο με αυτό μια πραγματικής κηλίδας όπως δίνεται από τη σχέση:

$$GT_{ij} = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{i^2+j^2}{2\sigma^2}\right)}, \quad i, j = 1, 2, \dots, 11, \quad (4.7)$$

Όπου σ είναι η τυπική απόκλιση, i είναι ο δείκτης της γραμμής των εικονοστοιχείων και j ο δείκτης των στηλών. Το σχήμα 4.11 απεικονίζει ένα 11X11 δυσδιάστατο πρότυπο όπως ακριβώς χρησιμοποιήθηκε για να εντοπίσει τις κηλίδες στην εικόνα.



Σχήμα 4.11: Δυσδιάστατο πρότυπο οι φωτεινότητες του οποίου ακολουθούν κανονική κατανομή.

Με αυτό το πρότυπο σαρώνεται παραθυρικά ολόκληρη η εικόνα και το εικονοστοιχείο, στο οποίο κεντροθετείται το πρότυπο κάθε φορά, χαρακτηρίζεται αν είναι ή όχι εικονοστοιχείο που ανήκει σε μια κηλίδα. Αυτό γίνεται υπολογίζοντας τον συντελεστή συσχέτισης του προτύπου αυτού με την 11X11 γειτονιά το εικονοστοιχείου στο οποίο είναι κεντροθετημένο το πρότυπο. Αυτό μας δίνει ένα μέτρο της ομοιότητας της γειτονιάς του εικονοστοιχείου με το πρότυπο. Αν συμβολίσουμε το πρότυπο με GT_{ij} και την γειτονιά της εικόνας με $W2D_{ij}$ ο συντελεστής συσχέτισης (correlation coefficient) δίνεται από την σχέση:

$$cc = \frac{\sum_i \sum_j (GT_{ij} - \overline{GT})(W2D_{ij} - \overline{W2D})}{\sqrt{(\sum_i \sum_j (GT_{ij} - \overline{GT})^2)(\sum_i \sum_j (W2D_{ij} - \overline{W2D})^2)}}, \quad (4.8)$$

όπου \overline{GT} είναι η μέση τιμή των φωτεινοτήτων του προτύπου, $\overline{W2D}$ είναι η μέση τιμή των φωτεινοτήτων του παραθύρου της εικόνας που είναι κεντροθετημένο το πρότυπο. Ο συντελεστής συσχέτισης παίρνει τιμές από -1 έως 1. Η τιμή -1 υποδηλώνει ότι το πρότυπο

με το παράθυρο της εικόνας δεν παρουσιάζουν καμία ομοιότητα, ενώ τιμή ίση με 1 σημαίνει ότι το παράθυρο και η εικόνα είναι απόλυτα ίδια. Για την ολοκλήρωση της μετατροπής της εικόνας σε δυαδική απαιτείται να οριστεί ένα κατώφλι για τον συντελεστή συσχέτισης πάνω από το οποίο το εικονοστοιχείο θα θεωρείται σήμα, ενώ κάτω από το οποίο το εικονοστοιχείο θα θεωρείται υπόβαθρο. Μετά από αρκετά πειράματα που έγιναν αυτό το κατώφλι για την τιμή του συντελεστή συσχέτισης ορίστηκε να είναι $cc = 0.3$.

Το βασικό πλεονέκτημα αυτής της τεχνικής έγκειται στο γεγονός ότι έχει την δυνατότητα να απαλείψει αντικείμενα της εικόνας που μπορεί να είναι φωτεινά, που δεν έχουν όμως σχήμα παρόμοιο με το σχήμα που πρέπει να έχει μια κηλίδα. Αντιθέτως, η τεχνική αυτή μειονεκτεί γιατί αφενός απαιτείται το μέγεθος του προτύπου, και αφετέρου γιατί απαιτείται μια τιμή κατωφλίωσης για τον συντελεστή συσχέτισης.

Η άλλη τεχνική που χρησιμοποιήθηκε για την αρχική κατάτμηση είναι η τεχνική του Ότσου, η οποία είναι μια αυτόματη τεχνική κατωφλίωσης. Η κατωφλίωση αυτή θα οδηγήσει στον διαχωρισμό των εικονοστοιχείων σε δύο κλάσεις. Έστω ότι L_1 και L_2 είναι αντίστοιχα η κλάση των εικονοστοιχείων με φωτεινότητα χαμηλότερη του κατωφλιού και η κλάση των εικονοστοιχείων που έχουν φωτεινότητα υψηλότερη του κατωφλιού. Αν το κατώφλι για παράδειγμα βρεθεί να είναι το επίπεδο το γκρι Th_2 , τότε οι δύο κλάσεις θα περιέχουν τα επίπεδα του γκρι:

$$\begin{cases} L_1 = \{0, 1, \dots, Th_2\} \\ L_2 = \{Th_2 + 1, Th_2 + 2, \dots, L - 1\} \end{cases} \quad (4.9)$$

όπου L είναι ο συνολικός αριθμός των επιπέδων του γκρι της εικόνας. Εάν σ_c^2 , σ_{Tot}^2 είναι η διακύμανση της κλάσης, και η συνολική διακύμανση αντίστοιχα, ένα βέλτιστο κατώφλι μπορεί να βρεθεί ελαχιστοποιώντας την παρακάτω εξίσωση σε σχέση με το Th_2 :

$$Opt = \frac{\sigma_c^2}{\sigma_{Tot}^2}, \quad (4.10)$$

Το βέλτιστο κατώφλι δηλαδή Th_2^{opt} θα δίνεται από την σχέση:

$$Th_2^{opt} = \arg \min_{Th_2} (Opt), \quad (4.11)$$

ισχύουν ότι:

$$\sigma_{Tot}^2 = \sum_{i=0}^{L-1} (i - \mu_T)^2 P_i, \quad \mu_T = \sum_{i=0}^{L-1} iP_i, \quad (4.12)$$

$$\sigma_c^2 = w_1 w_2 (\mu_1 \mu_2)^2, \quad w_1 = \sum_{i=0}^{Th_2} P_i, \quad w_2 = 1 - w_1, \quad (4.13)$$

$$\mu_2 = \frac{\mu_T - \mu_{Th_2}}{1 - \mu_1}, \quad \mu_1 = \frac{\mu_T}{w_1}, \quad \mu_{Th_2} = \sum_{i=0}^{Th_2} iP_i, \quad P_i = \frac{n_i}{n}, \quad (4.14)$$

όπου n_i είναι ο αριθμός των εικονοστοιχείων του i -οστού επιπέδου του γκρι και n είναι ο συνολικός αριθμός των εικονοστοιχείων ο οποίος ορίζεται ως:

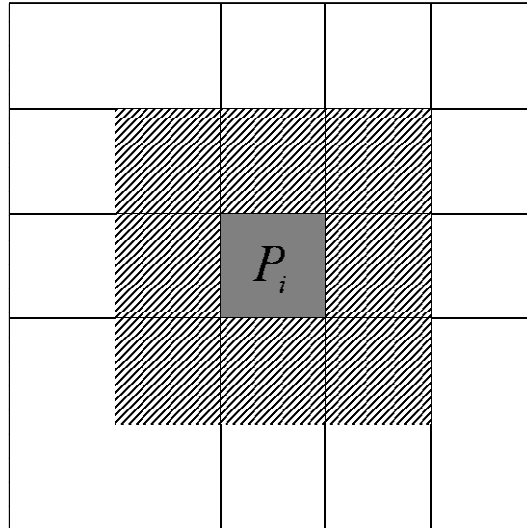
$$n = \sum_{i=0}^{L-1} n_i \quad (4.15)$$

Τέλος P_i είναι η πιθανότητα εμφάνισης του i -οστού επιπέδου του γκρι.

Σε σχέση με την τεχνική της σύμπτωσης προτύπων η τεχνική του Ότσου είναι σαφώς πιο γρήγορη και λιγότερο απαιτητική ως προς την υπολογιστική ισχύ. Επίσης δεν θεωρεί το μέγεθος της κάθε κηλίδας ότι είναι 10 περίπου εικονοστοιχεία ούτε ότι η κατανομή της φωτεινότητας των εικονοστοιχείων είναι κανονική. Επομένως η τεχνική αυτή είναι απαλλαγμένη από τις θεωρήσεις αυτές, οι οποίες πηγάζουν από τα χαρακτηριστικά που πρέπει θεωρητικά να έχει μια κηλίδα. Ωστόσο αυτό το γεγονός αποτελεί και μειονέκτημα της Ότσου σε σχέση με την τεχνική σύμπτωσης προτύπων διότι η τεχνική σύμπτωσης προτύπων καταφέρνει να κάνει μια πρώτη επιλογή των πραγματικών κηλίδων μέσα σε όλα τα φωτεινά αντικείμενα που εντοπίζονται και τα οποία μπορεί να περιέχουν πολλά τεχνουργήματα.

Για να εντοπιστούν και να σημειωθούν τα αντικείμενα της εικόνας δεν αρκεί να βρεθούν τα εικονοστοιχεία τα οποία ανήκουν σε φωτεινά αντικείμενα, αλλά να βρεθούν τα

συνδεδεμένα αντικείμενα. Για τον σκοπό αυτό εντοπίζονται τα συνδεδεμένα αντικείμενα 8 γειτόνων. (8-connected objects). Δύο δηλαδή εικονοστοιχεία θεωρούνται ότι βρίσκονται στο ίδιο αντικείμενο όταν αυτά γειτνιάζουν με συνδετικότητα 8-γειτονιάς (δηλαδή όταν έχουν τουλάχιστον μια κοινή κορυφή, όπως ακριβώς φαίνεται στο σχήμα 4.12. Το εικονοστοιχείο P_i στο παρακάτω σχήμα έχει συνδετικότητα 8-γειτονιάς με όλα τα γραμμοσκιασμένα εικονοστοιχεία.



Σχήμα 4.12: Συνδετικότητα 8-γειτονιάς

Το τελικό βήμα του εντοπισμού των κηλίδων που αντιστοιχούν σε υβριδοποιημένα δείγματα είναι υπολογισμός των κέντρων των αντικειμένων που έχουν εντοπιστεί. Τα κέντρα αυτά θα χρησιμοποιηθούν για να προσδιοριστούν οι θέσεις των κηλίδων χαμηλής φωτεινότητας στο επόμενο βήμα. Στην επεξεργασία μικροσυστοιχιών έχει επικρατήσει αντί να υπολογίζεται το καθαρά γεωμετρικό κέντρο της κάθε κηλίδας, να υπολογίζεται το κέντρο μάζας [110]. Στο κέντρο μάζας δεν συνεισφέρουν μόνο η συντεταγμένες των εικονοστοιχείων που αποτελούν το αντικείμενο, αλλά και η φωτεινότητα τους. Η φωτεινότητα του εικονοστοιχείου λειτουργεί σαν συντελεστής βαρύτητας στις συντεταγμένες των εικονοστοιχείων του αντικειμένου. Οι συντεταγμένες του κέντρου μάζας (X, Y) μιας κηλίδας δίνονται από της παρακάτω εξισώσεις:

$$X = \frac{\sum_{i \in Obj} x_i I_i}{\sum_{i \in Obj} I_i}, \quad Y = \frac{\sum_{i \in Obj} y_i I_i}{\sum_{i \in Obj} I_i}, \quad (4.16)$$

όπου x_i και y_i είναι οι συντεταγμένες του i -οστού εικονοστοιχείου, I_i είναι η φωτεινότητα του i -οστού εικονοστοιχείου, ενώ Obj είναι η λίστα από τα εικονοστοιχεία που ανήκουν στο συγκεκριμένο αντικείμενο. Γενικά για να αποφευχθεί η σύγχυση των συντεταγμένων ενός κέντρου μιας κηλίδας στην εικόνα, και των συντεταγμένων ενός εικονοστοιχείου στην εικόνα, όταν απευθυνόμαστε σε κέντρο θα χρησιμοποιούμε κεφαλαία γράμματα, ενώ όταν απευθυνόμαστε σε εικονοστοιχεία με πεζά.

Παρά το γεγονός ότι και ο εντοπισμός των φωτεινών κηλίδων μπορεί να ακολουθηθεί ως έχει είτε για τετραγωνικό είτε για εξαγωνικό πλέγμα, μια καινοτόμος μέθοδος για την απαλοιφή τεχνουργημάτων χρησιμοποιήθηκε ως μέρος της μεθοδολογίας αυτής για το τετραγωνικό μόνο πλέγμα. Η μέθοδος αυτή είναι βασισμένη στην τεχνική ανεύρεσης των εκτρόπων (outliers detection) [178]. Το τετραγωνικό πλέγμα ευνοεί μια τέτοιου είδους τεχνική διότι οι κηλίδες είναι τοποθετημένες σε γραμμές και στήλες. Η μέθοδος που προτείνεται επιδιώκει ουσιαστικά να απαλείψει τα τεχνουργήματα τα οποία βρίσκονται ενδιάμεσα στις γραμμές ή στις στήλες των κηλίδων, και τα οποία είναι πιθανότατα προϊόντα κάποιας σκόνης ή άλλης μόλυνσης πάνω στην μικροσυστοιχία κατά την διάρκεια της σάρωσης. Για τον σκοπό αυτό χρησιμοποιούνται οι συντεταγμένες των κέντρων μάζας που έχουν υπολογιστεί ανωτέρω. Πιο συγκεκριμένα αν τα κέντρα μάζας ενός αντικειμένου ικανοποιούν τις παρακάτω συνθήκες, τα αντικείμενα αυτά χαρακτηρίζονται ως έκτροπες και απαλείφονται:

$$X(p) < \text{mean}(COL_x) - 2 * \text{stddev}(COL_x), \quad (4.17)$$

$$X(p) > \text{mean}(COL_x) + 2 * \text{stddev}(COL_x), \quad (4.18)$$

$$Y(p) > \text{mean}(ROW_y) - 2 * \text{stddev}(ROW_y), \quad (4.19)$$

$$Y(p) > \text{mean}(ROW_y) + 2 * \text{stddev}(ROW_y), \quad (4.20)$$

Όπου X , Y είναι οι συντεταγμένες του κέντρου μάζας της κηλίδας p , COL_x είναι το διάνυσμα των τετμημένων x των κηλίδων υψηλής φωτεινότητας που ανήκουν στην στήλη στην οποία ανήκει και το p , και ROW_y είναι το διάνυσμα των τεταγμένων y των κηλίδων

υψηλής φωτεινότητας που ανήκουν στην γραμμή της συστάδας στην οποία ανήκει και το p .

4.4.Εντοπισμός των κηλίδων που δεν έχουν υβριδοποιηθεί

Μια αξιόπιστη μέθοδος για τον εντοπισμό των κηλίδων χαμηλής φωτεινότητας μπορεί να περιορίσει σε μεγάλο βαθμό τις «χαμένες τιμές» (missing values), οι οποίες παρουσιάζονται στα παραγόμενα δεδομένα. Η αντιμετώπιση των χαμένων τιμών αποτελεί άλλο ένα πεδίο έρευνας στην επεξεργασία των δεδομένων. Στο εδάφιο αυτό θα αναπτύξουμε την αρχική προσέγγιση [179-180] που υλοποιήθηκε για εικόνες τετραγωνικού πλέγματος, ενώ μετέπειτα θα περιγραφεί η γενικευμένη προσέγγιση που ακολουθήθηκε σε αυτό το βήμα [181-183].

4.4.1.Εντοπισμός μη υβριδοποιημένων κηλίδων σε εικόνες τετραγωνικού πλέγματος

Το βασικό χαρακτηριστικό του τετραγωνικού πλέγματος είναι το γεγονός ότι η κηλίδες είναι παραταγμένες στην εικόνα σε γραμμές και στήλες. Αυτό όπως φάνηκε και ανωτέρω με την εφαρμογή της τεχνικής της ανεύρεσης εκτρόπων, το τετραγωνικό πλέγμα δίνει την δυνατότητα της επεξεργασία των γραμμών και των στηλών των κηλίδες. Ο εντοπισμός των κηλίδων χαμηλής φωτεινότητας ξεκινά με τον εντοπισμό των τεσσάρων γωνιών (των κηλίδων στις τέσσερις γωνίες) της συστάδας. Αυτό είναι εύκολο να γίνει γιατί έχοντας εντοπίσει αρκετές κηλίδες από την συστάδα μπορούμε εύκολα να υποθέσουμε που περίπου θα βρίσκονται η τέσσερις «γωνιακές» κηλίδες. Για παράδειγμα, το κέντρο της κηλίδας στην επάνω αριστερή γωνία της συστάδας αναμένεται να έχει τετμημένη x περίπου ίση με τις τετμημένες των κέντρων, τα οποία ανήκουν στην πρώτη στήλη της συστάδας, ενώ η τεταγμένη y είναι ίση περίπου με τις τεταγμένες των κηλίδων που ανήκουν στην πρώτη γραμμή της συστάδας. Αν δεν έχει εντοπιστεί κανένα αντικείμενο κοντά στην γειτονιά αυτή τότε τοποθετείται ένα κέντρο κηλίδας με τετμημένη ίση με την μέση τιμή των τετμημένων των κηλίδων που ανήκουν στην πρώτη στήλη της συστάδας και τεταγμένη ίση με την μέση τιμή των τεταγμένων των κηλίδων που ανήκουν στην πρώτη γραμμή της συστάδας. Αυτό γίνεται για να αντιμετωπιστεί η περίπτωση κάποιες

από της γωνιακές κηλίδες να είναι κηλίδες χαμηλής φωτεινότητας που δεν έχουν εντοπιστεί από το προηγούμενο βήμα. Είναι σημαντικό να γίνει αυτό γιατί οι γωνιακές κηλίδες είναι σημεία αρχής και τέλος του αλγορίθμου.

Μια βασική ποσότητα που χαρακτηρίζει το πλέγμα των κηλίδων, η οποία θα χρησιμοποιηθεί κατά την διάρκεια της επεξεργασίας, είναι η επικρατούσα απόσταση μεταξύ γειτονικών κηλίδων (*mode_dist*) όπως αυτή ορίστηκε στο εδάφιο 4.2.2. Η προτεινόμενη μέθοδος για το τετραγωνικό πλέγμα είναι μια προσέγγιση κηλίδα με κηλίδα. Κατά την επεξεργασία διατρέχονται η γραμμές της συστάδας εξετάζοντας την απόσταση των διαδοχικών κηλίδων. Σε κάθε επανάληψη του αλγορίθμου εξετάζεται ένα ζεύγος διαδοχικών κηλίδων. Η μέθοδος προσθέτει μη υβριδοποιημένες κηλίδες στα σημεία που η απόσταση μεταξύ των δύο διαδοχικών εντοπισμένων κηλίδων είναι πολλαπλάσια από την επικρατούσα απόσταση. Όταν συμβαίνει αυτό, ο αριθμός των κηλίδων Q που θα προστεθούν μεταξύ δύο γειτονικών κηλίδων υψηλής φωτεινότητας (p και $p+1$) ορίζεται από τον παρακάτω τύπο:

$$Q = \left\lceil \frac{(S_x(p+1) - S_x(p))}{\text{mode}(d_{sx_p})} - 1 \right\rceil, \quad (4.21)$$

όπου S_x είναι το διάνυσμα των τετμημένων x των κηλίδων υψηλής φωτεινότητας, και d_{sx_p} είναι το διάνυσμα όλων των ευκλείδειων αποστάσεων. Οι συντεταγμένες των κέντρων των κηλίδων χαμηλής φωτεινότητας που θα προστεθούν στο πλέγμα NE_x και NE_y δίνονται από τις σχέσεις:

$$NE_x(q) = S_x(p) + q * (\text{median}(d_{sx_p})), \quad (4.22)$$

και

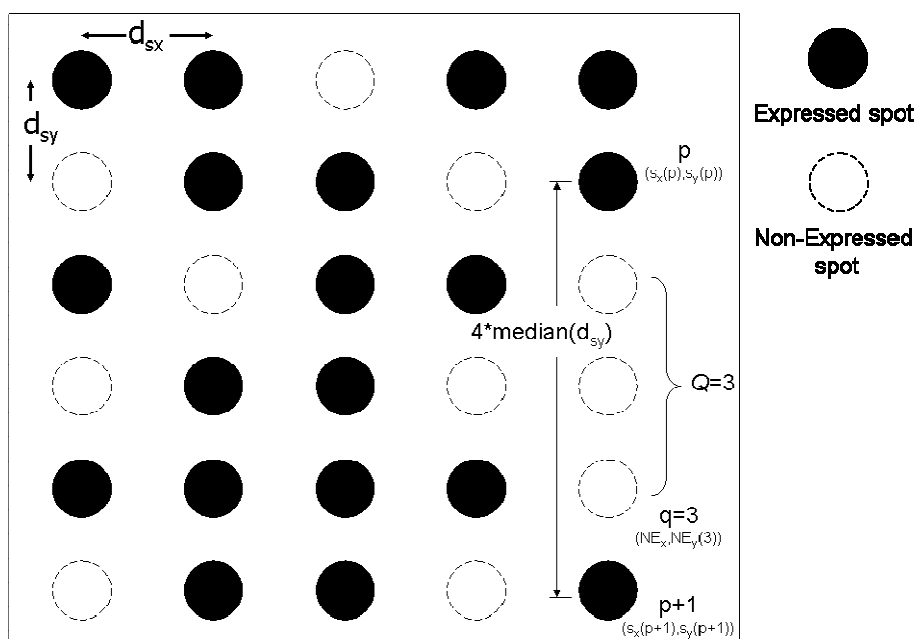
$$NE_y(q) = \text{mean}(ROW_y), \quad (4.23)$$

όπου ROW_y είναι το διάνυσμα των τεταγμένων y των κέντρων των κηλίδων υψηλής φωτεινότητας των κέντρων που ανήκουν στην γραμμή που ανήκουν και τα p και $p+1$.

Αφού η μέθοδος διατρέξει όλη την συστάδα, η διαδικασία της ανεύρεσης εκτρόπων, όπως αυτή περιγράφεται από τις εξισώσεις (4.17-4.20) επαναλαμβάνεται με σκοπό να απαλειφθούν κέντρα μη υβριδοποιημένων κηλίδων, τα οποία δεν έχουν τοποθετηθεί σωστά από την παραπάνω διαδικασία. Τελικά τα κέντρα τα οποία απαλείφονται αντικαθίστανται με κέντρα τα οποία έχουν τετμημένη την μέση τιμή των τετμημένων των κέντρων που ανήκουν στην ίδια στήλη της συστάδας και τεταγμένη την μέση τιμή των τεταγμένων των κέντρων που ανήκουν στην ίδια γραμμή της συστάδας.

Η παραπάνω διαδικασία δεν γίνεται μόνο διατρέχοντας τις γραμμές της συστάδας αλλά επαναλαμβάνεται και η ίδια με τις αντίστοιχες εξισώσεις και για τις στήλες από κηλίδες. Η προτεινόμενη μέθοδος για τον προσδιορισμό των κέντρων των κηλίδων χαμηλής φωτεινότητας παρουσιάζεται στο σχήμα 4.13 για την περίπτωση όπου προσδιορίζονται οι θέσεις των κέντρων διατρέχοντας μια στήλη.

Με αυτήν την διαδικασία ένα πλήρες πλέγμα έχει δημιουργηθεί που περιέχει όλα τα κέντρα που αντιστοιχούν στις κηλίδες που υπάρχουν στην συστάδα, και που είναι απαλλαγμένο από οποιαδήποτε τεχνούργημα.



Σχήμα 4.13: Η διαδικασία εντοπισμού μη υβριδοποιημένων κηλίδων για εικόνες τετραγωνικού πλέγματος.

4.4.2.Γενικευμένη αλγόριθμος εντοπισμού των μη υβριδοποιημένων κηλίδων

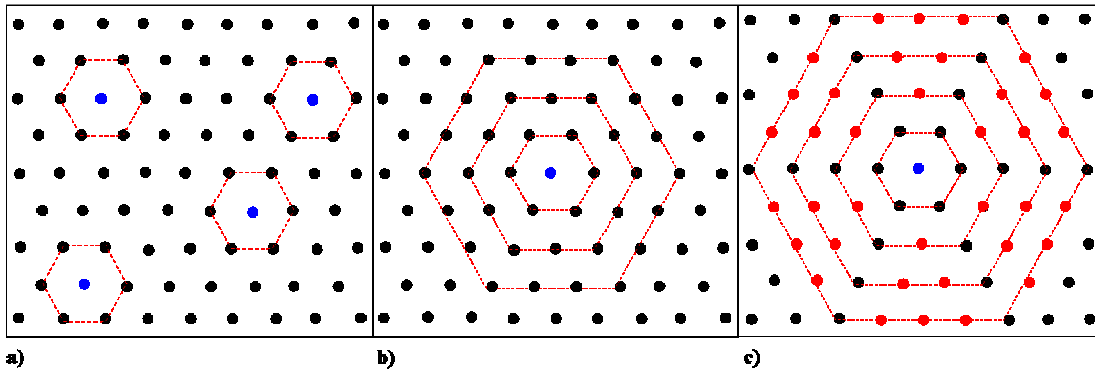
Η παραπάνω προσέγγιση για τον εντοπισμό των θέσεων των σκοτεινών κηλίδων για το τετραγωνικό πλέγμα είναι αναμφίβολα βασισμένη στις μεθόδους τις οποίες ονομάσαμε στην ανασκόπηση της βιβλιογραφίας προσεγγίσεις κηλίδα-με-κηλίδα. Η θέση κάθε κηλίδας χαμηλής φωτεινότητας προσδιορίζεται από τις θέσεις των γειτόνων της, ανατρέχοντας σε κάθε μια κηλίδα ξεχωριστά. Όπως ήδη αναφέρθηκε στο εδάφιο 4.1 ο βασικός λόγος για τον οποίο χρησιμοποιήθηκε το εξαγωνικό πλέγμα αντί του τετραγωνικού ήταν καθαρά η οικονομία χώρου. Αναπόφευκτα λοιπόν, η εικόνα μιας μικροσυστοιχίας που έχει χρησιμοποιήσει εξαγωνικό πλέγμα για την χωροθέτηση των κηλίδων της θα περιέχει έναν πολύ μεγάλο αριθμό από κηλίδες σε σχέση με την εικόνα μιας μικροσυστοιχίας που χρησιμοποιεί τετραγωνικό πλέγμα. Ενδεικτικά αναφέρουμε ότι το σύνολο δεδομένων με τετραγωνικό πλέγμα που χρησιμοποιήσαμε και το οποίο λάβαμε από την βάση δεδομένων του Στάνφορντ [74], περιέχει περίπου 600 κηλίδες ανά συστάδα, ενώ οι εικόνες από τα πειράματα της Illumina [184] περίπου 150000. Το τεράστιο πλήθος των κηλίδων απαιτεί έναν πιο έξυπνο αλγόριθμο για την επεξεργασία της κάθε συστάδας.

Ο γενικευμένος αλγόριθμος ο οποίος αναπτύχθηκε για τον εντοπισμό των μη υβριδοποιημένων κηλίδων σκοπεύει στην αξιοποίηση των ιδιοτήτων των δύο πλεγμάτων. Σκοπός είναι ο εντοπισμός να μην γίνεται κηλίδα με κηλίδα, αλλά οι θέσεις των κηλίδων της εικόνας να εξάγεται άμεσα από τις κορυφές του πλέγματος. Ο αλγόριθμος αυτός ονομάστηκε αλγόριθμος των αναπτυσσόμενων ομόκεντρων πολυγώνων (Growing Concentric Polygons – GCP). Τα πολύγωνα τα οποία αναπτύσσονται αναφέρονται είτε σε τετράγωνα όταν η εικόνα χρησιμοποιεί τετραγωνικό πλέγμα για την χωροθέτηση των κηλίδων, είτε σε εξάγωνα όταν κάνει χρήση εξαγωνικού πλέγματος. Τρεις βασικές ιδιότητες που ισχύουν και στους δύο τύπους πλεγμάτων χρησιμοποιήθηκαν για την υλοποίηση του αλγορίθμου:

- Κάθε κηλίδα στην εικόνα αποτελεί κέντρο ενός πολυγώνου.
- Κάθε κηλίδα στην εικόνα αποτελεί κέντρο ομόκεντρων πολυγώνων.

- Ο αριθμός των κηλίδων τα οποία είναι τοποθετημένα στο περίγραμμα των ομόκεντρων πολυγώνων είναι εύκολα υπολογίσιμος.

Στο σχήμα 4.14 αναπαρίστανται οι τρεις αυτές ιδιότητες για το εξαγωνικό πλέγμα.



Σχήμα 4.14: Οι τρεις ιδιότητες του πλέγματος που χρησιμοποιήσαμε για την ανάπτυξη του αλγορίθμου των αναπτυσσόμενων ομόκεντρων πολυγώνων. α) οποιαδήποτε κηλίδα είναι κέντρο πολυγώνου, β) οποιαδήποτε κηλίδα είναι κέντρο ομόκεντρων πολυγώνων γ) ο υπολογισμός των κηλίδων στο περίγραμμα το κάθε πολυγώνου μπορεί να υπολογιστεί.

| | |
|---|--|
| <p>Ομόκεντρα Τετράγωνα:</p> <ul style="list-style-type: none"> ▪ 1^ο: 4 κηλίδες είναι τοποθετημένες στις γωνίες και από μια κηλίδα στο μέσο κάθε ακμής = 8 ▪ 2^ο: 4 κηλίδες στις γωνίες και από 3 κηλίδες στις ακμές = 16 ▪ 3^ο: 4 κηλίδες στις γωνίες και από 5 κηλίδες στις ακμές = 24 • • • ▪ n-οστό: 4 κηλίδες στις γωνίες, 4 στο μέσο της κάθε ακμής από άλλες $2*(n-1)$ παρεμβάλλονται σε κάθε ακμή <p style="text-align: center;">$n = 8 * level$</p> | <p>Ομόκεντρα Εξάγωνα:</p> <ul style="list-style-type: none"> ▪ 1^ο: 6 κηλίδες τοποθετημένες όλες στις 6 γωνίες = 6 ▪ 2^ο: Σε κάθε ακμή του εξαγώνου παρεμβάλλεται μια κηλίδα = 12 ▪ 3^ο: Σε κάθε ακμή παρεμβάλλονται δύο κηλίδες = 18 • • • ▪ n-οστό: Σε κάθε ακμή παρεμβάλλονται $n-1 spots$ <p style="text-align: center;">$n = 6 * level$</p> |
|---|--|

Σχήμα 4.15: Τα επίπεδα των ομόκεντρων πολυγώνων και ο αριθμός των κηλίδων επάνω στο περίγραμμά τους.

Ο αριθμός των κηλίδων N_i που βρίσκονται επάνω στο περίγραμμά του i -οστού πολύγωνα εξαρτάται μόνο από το επίπεδο του ομόκεντρου πολυγώνου, και διαφοροποιείται βέβαια ανάλογα με το πολύγωνα που χρησιμοποιείται ως δομική μονάδα (τετράγωνο ή εξάγωνο). Αυτό γίνεται εύκολα κατανοητό αν κοιτάξουμε αναδρομικά τα επίπεδα των πολυγώνων

και τις κηλίδες επάνω στο περίγραμμα τους όπως φαίνονται στο σχήμα 4.14γ. Αναδρομικά λοιπόν μπορούν να εξαχθούν τα συμπεράσματα του σχήματος 4.15 και για τα αναπτυσσόμενα τετράγωνα και για τα αναπτυσσόμενα εξάγωνα.

Οι αριθμοί των κηλίδων επάνω στο περίγραμμα του κάθε πολυγώνου υπολογίζονται:

$$N_l^{rec} = 8 \cdot l, \quad (4.24)$$

για τα ομόκεντρα τετράγωνα και:

$$N_l^{hex} = 6 \cdot l, \quad (4.25)$$

όπου l είναι το επίπεδο του ομόκεντρου πολυγώνου. Για το αριθμό δύο διαδοχικών πολυγώνων ισχύει επίσης ότι:

$$N_l^{rec} - N_{l-i}^{rec} = 8, \quad (4.26)$$

και

$$N_l^{hex} - N_{l-i}^{hex} = 6. \quad (4.27)$$

Ο Αλγόριθμος των αναπτυσσόμενων ομόκεντρων πολυγώνων περιγράφεται συνοπτικά από τον ψευδοκώδικα του Σχήματος 4.16.

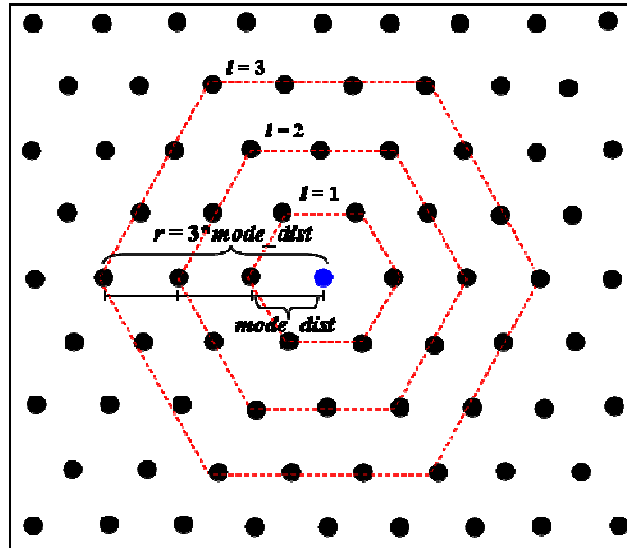
```

Τυχαία επιλογή μιας φωτεινής κηλίδας
r = mode_dist
FOR l = 1:size(block) //l είναι το επίπεδο του πολυγώνου
    IF grid = hex
        ni = 6*l
    ELSE grid=rec
        ni = 8*l
    ENDF
    //εξέταση των θέσεων των κηλίδων στο περίγραμμα του l-οστού πολυγώνου
    FOR n = 1:ni
        IF (Υπάρχει ήδη εν.ωι.ομένη κηλίδα στην 3x3 γειτονιά της
            παραγόμενης κηλίδας)
            THEN
                Η παραγόμενη κηλίδα απαλείφεται και παραμένει η ήδη
                ενισπιόμενη κηλίδα
            END
        END
        r = r + mode_dist
    END //επόμενο πολύγωνο

```

Σχήμα 4.16: Ψευδοκώδικας του αλγόριθμου των αναπτυσσόμενων ομόκεντρων πολυγώνων.

Ο αλγόριθμος ξεκινάει με την επιλογή μιας κηλίδας, η οποία έχει ήδη εντοπιστεί από το βήμα του εντοπισμού των υβριδοποιημένων κηλίδων. Η επικρατούσα απόσταση δύο γειτονικών κηλίδων είναι η ακτίνα του πολυγώνου πρώτου επιπέδου (όταν αναφερόμαστε σε ακτίνα πολυγώνου εννοούμε την ακτίνα του περιγεγραμμένου κύκλου του πολυγώνου), καθώς επίσης και προσαύξηση της ακτίνας των ομόκεντρων πολυγώνων, όπως φαίνεται και στο σχήμα 4.17, για τα αναπτυσσόμενα εξάγωνα.



Σχήμα 4.17: Η σχέση της επικρατούσας τιμής των αποστάσεων δύο γειτονικών κηλίδων με τις ακτίνες των ομόκεντρων εξάγωνων.

Η ακτίνα σε κάθε επανάληψη του αλγορίθμου δίνεται από σχέση:

$$r = \text{mod_dist} * l, \quad (4.28)$$

Σε κάθε μια επανάληψη του αλγορίθμου όπου το επίπεδο των πολυγώνων αυξάνεται κατά ένα, η ακτίνα του ομόκεντρου πολυγώνου αυξάνεται τόσο όσο είναι η επικρατούσα απόσταση. Όπως φαίνεται και στον ψευδοκώδικα ο αλγόριθμός επιλέγει τυχαία ένα ήδη εντοπισμένο αντικείμενο ως σημείο έναρξης του αλγορίθμου και το οποίο κατ' επέκταση θα είναι το κέντρο των ομόκεντρων πολυγώνων. Με κέντρο το σημείο αυτό και ακτίνα πολυγώνου αρχικοποιημένα ίση με την επικρατούσα τιμή των αποστάσεων των γειτονικών κηλίδων ξεκινούν οι επαναλήψεις του αλγορίθμου. Σε κάθε επανάληψη του αλγορίθμου αυξάνεται κατά ένα το επίπεδο των πολυγώνων. Η κηλίδες τοποθετούνται πάνω στο περίγραμμα του τρέχοντος πολυγώνου, ενώ ο αριθμός τους καθορίζεται από τις εξισώσεις (4.24-4.25). Τέλος, καθεμία από της παραγόμενες κηλίδες επαληθεύονται με τις κηλίδες

που έχουν ήδη βρεθεί. Για καθεμία παραγόμενη κηλίδα εξετάζεται η 3X3 γειτονιά της, με σκοπό αν υπάρχει ένα ήδη εντοπισμένο αντικείμενο σε αυτήν, η παραγόμενη κηλίδα απαλείφεται και παραμένει η ήδη εντοπισμένη.

Για να αυξηθεί η ευστάθεια του αλγόριθμου συμπεριλήφθηκαν τρεις βελτιώσεις για να αντιμετωπίσουν προβλήματα, τα οποία παρουσιάστηκαν κατά την εφαρμογή του. Όσο περνούν οι επαναλήψεις του αλγόριθμου και το επίπεδο των πολυγώνων αυξάνεται, η επικρατούσα απόσταση σταματάει να προσομοιώνει με ακρίβεια την προσαύξηση της ακτίνας των ομόκεντρων πολυγώνων. Αυτό συμβαίνει λόγω του γεγονότος ότι η εικόνα είναι ένα ψηφιοποιημένο σήμα και το κάθε εικονοστοιχείο της αντιστοιχίζεται με μια πραγματική απόσταση κατά την ψηφιοποίηση της. Για να γίνει κατανοητό πως αυτό επηρεάζει τον αλγόριθμο ας θεωρήσουμε την ιδανική περίπτωση το κάθε εικονοστοιχείο της εικόνας να αντιστοιχεί ακριβώς σε 1 μm πραγματικής απόστασης. Ας υποθέσουμε επίσης ότι η πραγματική απόσταση μεταξύ δύο κηλίδων είναι 10 μm και ότι η επικρατούσα τιμή που υπολογίστηκε στο εδάφιο 4.2.2 είναι 10 εικονοστοιχεία. Σε αυτήν την περίπτωση ξεκινώντας από ένα τυχαίο κέντρο, το πολύγωνο του πρώτου επιπέδου θα είχε ακτίνα 10 εικονοστοιχεία που θα αντιστοιχούσαν σε 10 μm, το δεύτερο 20 εικονοστοιχεία που θα αντιστοιχούσαν σε 20 μm, κ.ο.κ. Ας θεωρήσουμε τώρα την περίπτωση κατά την οποία το ένα εικονοστοιχείο δεν αντιστοιχεί σε ακριβώς 1 μm αλλά σε 1.04 μm. Ας υποθέσουμε επίσης ότι η πραγματική απόσταση μεταξύ δύο γειτονικών παραμένει 10 μm. Ο αλγόριθμος λόγω στρογγυλοποίησης θα υπολογίσει την επικρατούσα τιμή των αποστάσεων πάλι 10 εικονοστοιχεία δεδομένου ότι:

$$\begin{cases} 1.04 \cdot 9 = 9.36 \\ 1.04 \cdot 10 = 10.4 \\ 1.04 \cdot 11 = 11.44 \end{cases} \quad (4.29)$$

Άρα τα 10 εικονοστοιχεία εξακολουθούν να προσομοιώνουν καλύτερα την απόσταση μεταξύ δύο γειτονικών κηλίδων. Ωστόσο στο τρίτου επιπέδου πολύγωνο που θα παραχθεί μετά από δύο επαναλήψεις ο αλγόριθμος θα προσομοιώσει την ακτίνα του εξαγώνου αυτού με 30 εικονοστοιχεία ενώ θα έπρεπε να προσομοιωθούν με $30/1.04=28.84$ δηλαδή με 29 εικονοστοιχεία. Αυτά τα λάθη στρογγυλοποίησης συσσωρεύονται με τις πολλές επαναλήψεις και μετά από ένα αριθμό επαναλήψεων τα πολλαπλάσια της επικρατούσας

τιμής των αποστάσεων των γειτονικών κηλίδων σταματούν να προσομοιώνουν τις ακτίνες των ομόκεντρων πολυγώνων.

Για να αντιμετωπιστεί το πρόβλημα που προκύπτει από τις στρογγυλοποιήσεις, σε κάθε επανάληψη δεν εξετάζονται αυστηρά τα πολλαπλάσια της επικρατούσας τιμής αλλά και τιμές για την ακτίνα του εκάστοτε εξαγώνου ίσες με $(r-2, r-1, r+1, r+2)$, οι οποίες λαμβάνουν ένα σκορ ανάλογα με τον αριθμό των κέντρων που βρίσκονται πάνω στο περίγραμμα τους. Σε κάθε επανάληψη δηλαδή πέντε διαφορετικά ομόκεντρα πολύγωνα λαμβάνουν τιμή σύμφωνα με τον τύπο:

$$S_{rad} = \frac{N_{correct}}{N_i} \cdot 100\%, \quad (4.30)$$

όπου $N_{correct}$ είναι ο αριθμός των κηλίδων που έχουν πέσει πάνω στο περίγραμμα του πολυγώνου, N_i είναι ο συνολικός αριθμός των κηλίδων που έχουν δημιουργηθεί για το i -οστό πολύγωνο. Το πολύγωνο που επιλέγεται τελικά για να τοποθετηθούν στο περίγραμμα του τα κέντρα των παραγόμενων κηλίδων είναι αυτό που έχει την μεγαλύτερη τιμή S_{rad} .

Εφόσον η σωστή ακτίνα του τρέχοντος πολυγώνου δεν είναι η ακτίνα r που έχει καθοριστεί εξ αρχής από την εξίσωση 4.29 αλλά μια εκ των $(r-2, r-1, r+1, r+2)$, αυτό θα πρέπει να συνυπολογιστεί για την επόμενη επανάληψη του αλγορίθμου. Έτσι στο τέλος της συγκεκριμένης επανάληψης διορθώνεται η τιμή της ακτίνας που θα χρησιμοποιηθεί για την επόμενη επανάληψη. Αν για παράδειγμα το πολύγωνο που θα επιλεγεί είναι το εξαγώνο με ακτίνα $r+1$, η ακτίνα με την οποία θα λειτουργήσει η επόμενη επανάληψη θα είναι ίση με:

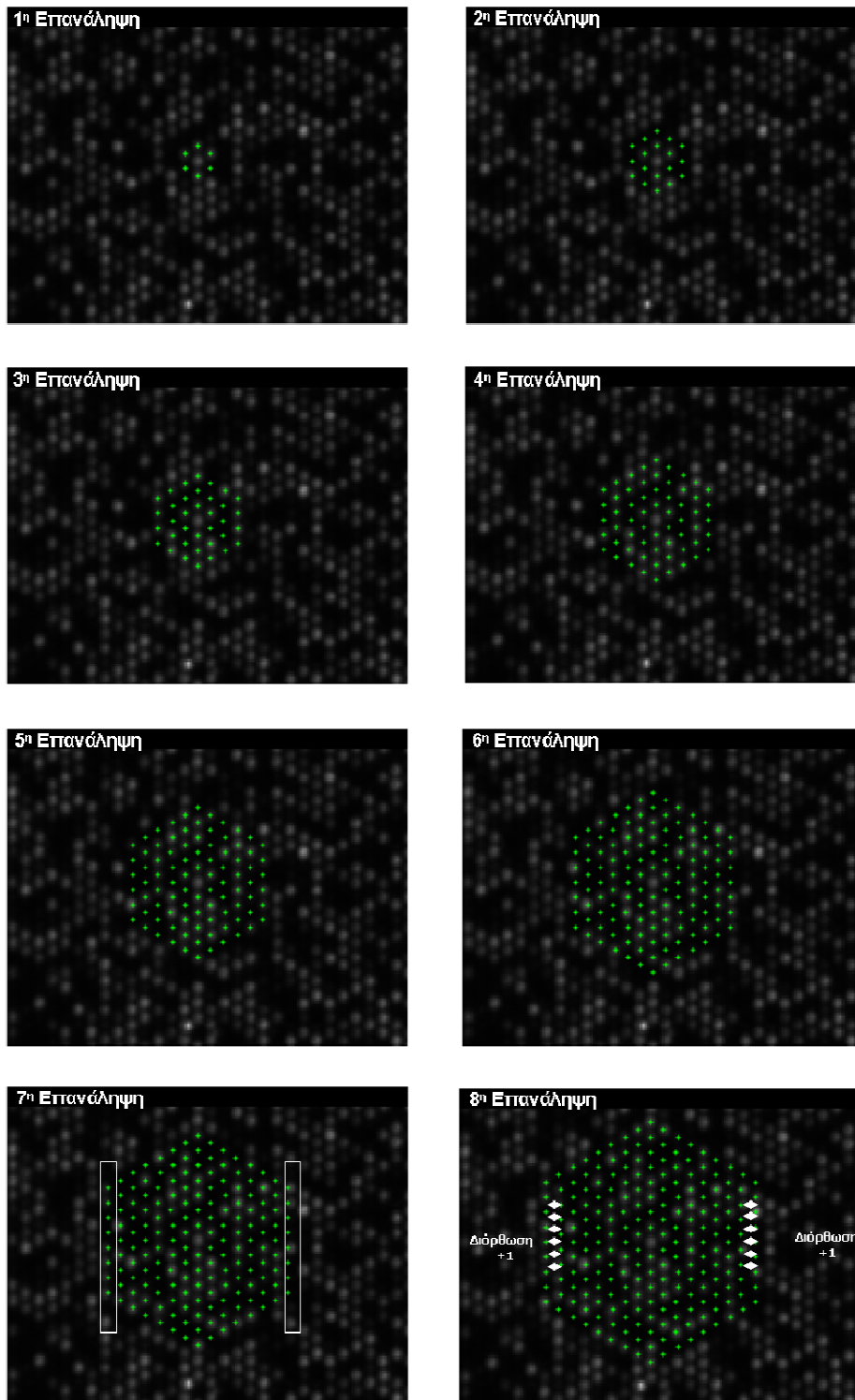
$$r_i = r_{i-1} + mod_dist + 1, \quad (4.31)$$

Όπου r_i είναι η ακτίνα της επανάληψης που ξεκινάει τώρα, ενώ r_{i-1} είναι η ακτίνα με την οποία ξεκίνησε η προηγούμενη επανάληψη.

Ο τρόπος με τον οποίο λειτουργεί η διόρθωση σε μια εικόνα με εξαγωνικό πλέγμα φαίνεται στο σχήμα 4.18. Μέχρι και την 6η επανάληψη φαίνεται τα κέντρα των παραγόμενων κηλίδων είναι σωστά τοποθετημένα. Ήδη από την 7η επανάληψη όμως τα συσσωρευμένα λάθη στρογγυλοποίηση έχουν αρχίσει να επιδρούν αρνητικά στον αλγόριθμο. Περισσότερο αισθητό είναι αυτό στις δύο κατακόρυφες ακμές του 7ου εξαγώνου τις οποίες έχουμε επισημάνει στο σχήμα με δύο ορθογώνια πλαίσια. Παρατηρούμε ότι τα κέντρα σε αυτές τις ακμές έχουν τοποθετηθεί μεν μέσα στα όρια των κηλίδων αλλά δείχνουν ότι τείνουν να ξεφύγουν από αυτές. Πράγματι στην 8η επανάληψη παρατηρούμε ότι έχει ενεργοποιηθεί η εν λόγω διόρθωση που περιγράψαμε ανωτέρω, διορθώνοντας την ακτίνα του 8ου εξαγώνου κατά ένα εικονοστοιχείο περισσότερο (διόρθωση +1). Στην διορθωμένη αυτή τιμή της ακτίνας θα προστεθεί στην επόμενη 9η επανάληψη η επικρατούσα τιμή ώστε να προκύψει η ακτίνα του 9ου εξαγώνου. Πέντε ακτίνες για το 9ο εξαγώνο θα εξεταστούν ξανά κ.ο.κ..

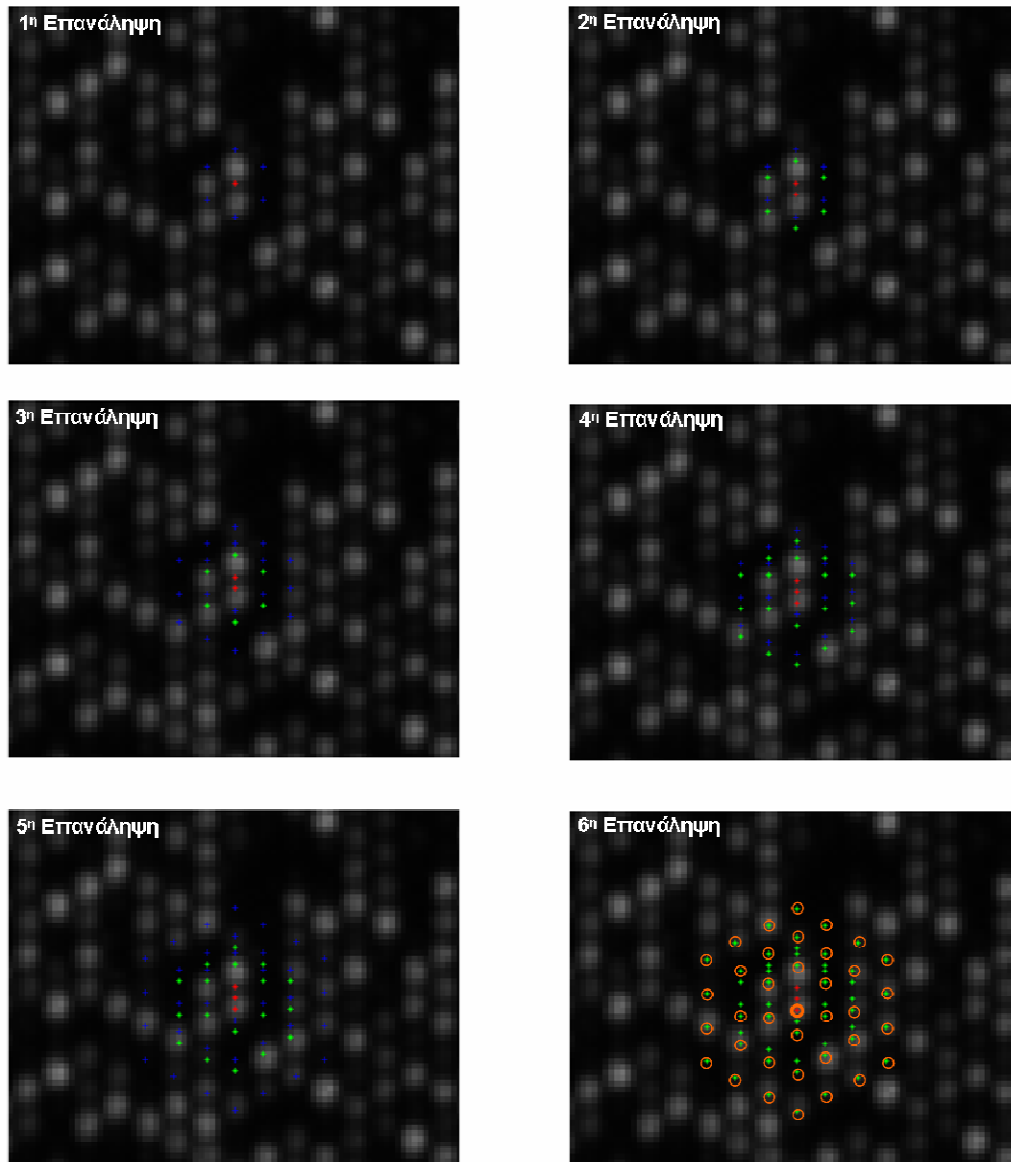
Η δεύτερη βελτίωση που έχει γίνει στο αλγόριθμο αφορά στο σημείο εκκίνησης του αλγορίθμου, το οποίο όπως έχουμε αναφέρει ότι επιλέγεται τυχαία από τα φωτεινά αντικείμενα που έχουν ήδη εντοπιστεί. Ένα τέτοιο αντικείμενο έχει πάρα πολλές πιθανότητες να είναι κάποια κηλίδα της εικόνας, δεν μπορεί όμως να αποκλειστεί η περίπτωση να είναι κάποιο τεχνούργημα. Επίσης υπάρχει μια ακόμα περίπτωση που μπορεί να δημιουργήσει πρόβλημα στην τυχαία επιλογή του σημείου εκκίνησης. Μερικές φορές δύο γειτονικές κηλίδες είναι πολύ φωτεινές με αποτέλεσμα να επηρεάζουν και τα υπόλοιπα εικονοστοιχεία της γύρω περιοχής (που κανονικά είναι εικονοστοιχεία υποβάθρου) αυξάνοντας τις φωτεινότητες και αυτών σε υψηλότερα επίπεδα από ένα τυπικό εικονοστοιχείο υποβάθρου.

Έτσι κατά τον εντοπισμό των υβριδοποιημένων κηλίδων μπορεί να εντοπιστούν οι συγκεκριμένες δύο κηλίδες ως ένα αντικείμενο. Όταν υπολογιστούν τα κέντρα μάζας του κάθε εντοπισμένου αντικειμένου το κέντρο μάζας του συσσωματώματος των δύο κηλίδων θα υπολογιστεί να είναι κάπου ανάμεσα στις δύο πολύ φωτεινές κηλίδες. Τέτοια συσσωματώματα κηλίδων που περιλαμβάνουν παραπάνω από μια κηλίδα μπορεί να δημιουργηθούν και στην περίπτωση που κατά την διάρκεια του βιολογικού πειράματος χυθεί φθορίζον δείγμα κάπου πάνω στην μικροσυστοιχία.



Σχήμα 4.18: Η διόρθωση των λαθών στρογγυλοποίησης

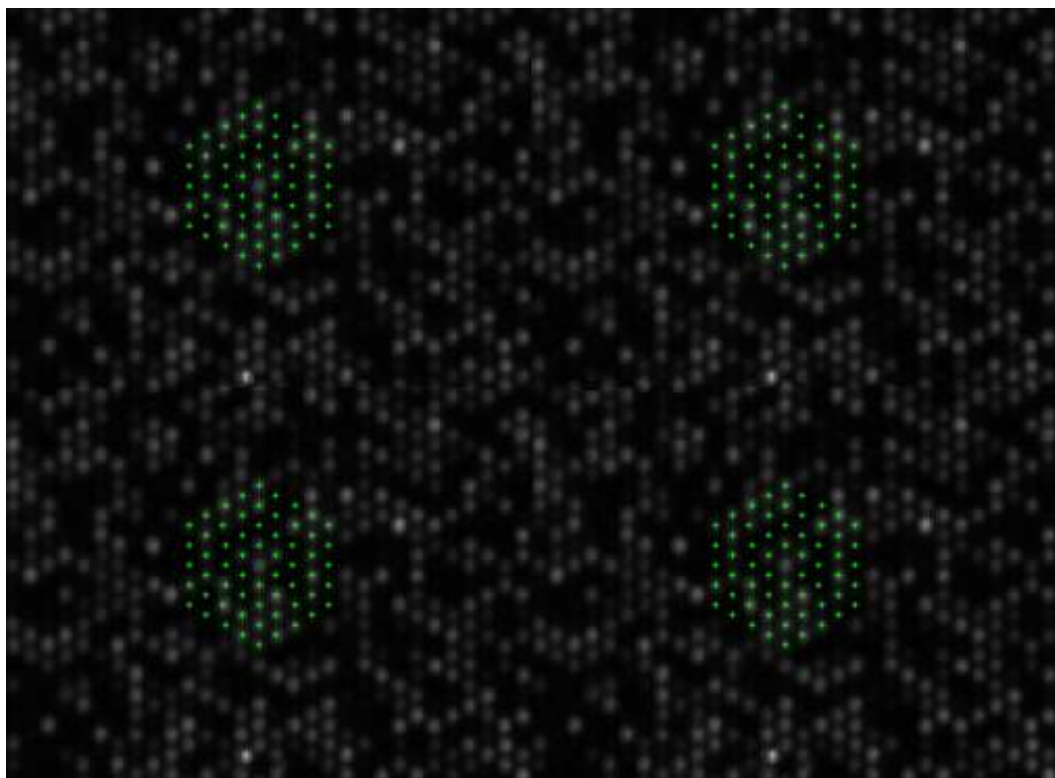
Η λάθος επιλογή του σημείου έναρξης θα μπορούσε να οδηγήσει σε απόλυτη αποτυχία του αλγορίθμου. Για να αποφευχθεί αυτός ο κίνδυνος η προτεινόμενη μέθοδος βελτιστοποιεί τις θέσεις των ήδη παραγόμενων κηλίδων σε κάθε επανάληψη του αλγορίθμου. Πιο συγκεκριμένα, κατά την διάρκεια της ι-οστής επανάληψης όλες οι κηλίδες που έχουν παραχθεί από το ξεκίνημα του αλγορίθμου έως και την ι-οστή επανάληψη θεωρούνται ως ένα άκαμπτο πλέγμα. Δεν επιτρέπονται δηλαδή σχετικές κινήσεις μεταξύ των κέντρων από το οποίο αποτελείται. Για το πλέγμα αυτών των σημείων υπολογίζεται το άθροισμα των φωτεινότητων που έχουν τα συγκεκριμένα σημεία στην αρχική εικόνα. Δεδομένου ότι τα κέντρα αυτά πρέπει να είναι κέντρα κηλίδων, πρέπει προφανώς τα περισσότερα να έχουν υψηλή φωτεινότητα (εκτός αν είναι κέντρα σκοτεινών κηλίδων). Όλο αυτό το πλέγμα κατά την διάρκεια μιας επανάληψης έχει την δυνατότητα να μετακινηθεί κατά δύο εικονοστοιχεία είτε πάνω είτε κάτω, είτε αριστερά, είτε δεξιά, είτε στις διαγώνιες διευθύνσεις. Η κατεύθυνση που θα μετακινηθεί το πλέγμα των σημείων μεγιστοποιεί το άθροισμα των φωτεινότητων των σημείων στην αρχική εικόνα. Όσο υψηλότερο είναι το άθροισμα των φωτεινότητων τόσο πιθανότερο είναι τα κέντρα του πλέγματος να είναι κέντρα κηλίδων. Το πλέγμα αυτό μπορούμε να πούμε ότι ακολουθεί ένα μικρό μονοπάτι στην εικόνα το οποίο μεγιστοποιεί την φωτεινότητα των σημείων του στην αρχική εικόνα. Δύο κινήσεις ανά επανάληψη είναι αρκετές εφόσον αυτό μπορεί να γίνει σε όλες τις επαναλήψεις. Εάν δεν υπήρχε ένα όριο στον αριθμό των εικονοστοιχείων που μπορεί να κινηθεί το πλέγμα μπορεί να δημιουργούταν πρόβλημα στις πρώτες επαναλήψεις, εφόσον στις πρώτες επαναλήψεις τα παραγόμενα κέντρα είναι λίγα σε αριθμό και έτσι δεν υπάρχει μεγάλη αξιοπιστία όσον αφορά στο άθροισμα των φωτεινότητων τους. Όσο ο αλγόριθμος προχωράει σε μεγαλύτερου επιπέδου εξάγωνα η συνεισφορά της συγκεκριμένης βελτίωσης σχεδόν εξαλείφεται εφόσον το πλέγμα έχει τοποθετηθεί σωστά από τις πρώτες επαναλήψεις. Στο σχήμα 4.19 φαίνεται από την πρώτη επανάληψη ότι το κέντρο που τυχαία έχει επιλεγεί για την ανάπτυξη του αλγορίθμου δεν ανταποκρίνεται σε ένα πραγματικό κέντρο κηλίδας, διότι οι δύο κηλίδες στις οποίες βρίσκεται ενδιάμεσα έχουν εντοπιστεί ως ένα ενιαίο αντικείμενο (συσσωμάτωμα). Ωστόσο παρατηρούμε ότι από την 4η επανάληψη το κέντρο αυτό έχει μετακινηθεί και έχει βρει την κηλίδα που στην αρχή βρισκόταν κάτω από το κέντρο. Μαζί με το σημείο εκκίνησης έχουν μετακινηθεί και όλα τα υπόλοιπα κέντρα τα οποία είχαν μέχρι εκείνη την στιγμή παραχθεί.



Σχήμα 4.19: Βελτιστοποίηση της θέσης του σημείου εκκίνησης του αλγορίθμου.

Τέλος η 3η βελτίωση του αλγορίθμου έχει στόχο την αποφυγή προβλημάτων στρέβλωσης στην εικόνα. Παρά τις δύο παραπάνω βελτιώσεις όταν ο αλγόριθμος φτάνει σε πολύ μεγάλο επίπεδο πολύγωνα αδυνατεί να εντοπίσει με ακρίβεια (στην 3X3 περιοχή) τις κηλίδες πάνω στο περίγραμμα των πολυγώνων λόγω της στρέβλωσης της εικόνας. Είναι σημαντικό να αναφέρουμε ότι σε μια συστάδα ο αλγόριθμος πραγματοποιεί εκατοντάδες επαναλήψεις, και άρα παράγει εκατοντάδες ομόκεντρα πολύγωνα. Για να αντιμετωπιστεί το παραπάνω πρόβλημα ο αλγόριθμος δεν ξεκινάει από ένα μόνο σημείο εκκίνησης αλλά από περισσότερα. Στην αρχή δηλαδή ο αλγόριθμος επιλέγει τυχαία ένα σύνολο σημείων

ώστε το καθένα να αναπτυχθεί μόνο σε μια περιοχή της εικόνας. Με τον τρόπο αυτό τα μέγιστα επίπεδα των ομόκεντρων πολυγώνων θα μειωθούν σε μεγάλο βαθμό. Στο παρακάτω σχήμα 4.20 φαίνεται μια περιοχή της εικόνας στην οποία αναπτύσσονται ομόκεντρα εξάγωνα από τέσσερα διαφορετικά σημεία εκκίνησης.



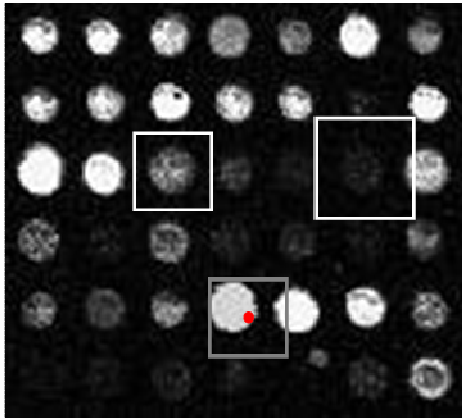
Σχήμα 4.20: Τέσσερα σημεία εκκίνησης του αλγορίθμου.

Ο βέλτιστος αριθμός των σημείων εκκίνησης είναι και αυτό ένα θέμα περαιτέρω μελέτης ώστε να επιτευχθούν τα καλύτερα αποτελέσματα.

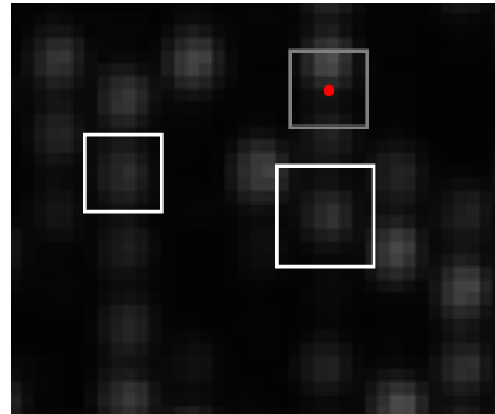
4.5. Τοποθέτηση πλέγματος

Η τοποθέτηση του πλέγματος είναι μια διαδικασία, η οποία στις περισσότερες εργασίες στην βιβλιογραφία δεν γίνεται με ευφυή τρόπο [75, 92, 97, 110]. Αφού εντοπιστούν τα κέντρα των κηλίδων οι μέθοδοι απλώς τοποθετούν μια τετράγωνη περιοχή γύρω από αυτά, με καθορισμένο μέγεθος. Όπως παρατηρείται και στο παρακάτω σχήμα 4.21 μια τέτοια τετραγωνική περιοχή μπορεί να περιέχει την ζητούμενη κηλίδα αλλά ανάλογα με το μέγεθος του τετραγώνου μπορεί να περιέχει και εικονοστοιχεία από γειτονικές κηλίδες. Το φαινόμενο αυτό μπορεί να γίνει πολύ πιο αισθητό στο εξαγωνικό πλέγμα όπου γύρω από

την ζητούμενη κηλίδα υπάρχουν έξι γειτονικές κηλίδες σε πολύ κοντινή απόσταση. Επίσης το φαινόμενο μπορεί να γίνει ακόμα πιο έντονο όταν το κέντρο της κηλίδας δεν έχει εντοπιστεί με μεγάλη ακρίβεια. Σε αυτή την περίπτωση η πιθανότητα να υπάρχουν στην τετράγωνη περιοχή εικονοστοιχεία σήματος από γειτονικές κηλίδες είναι πολύ μεγάλη, όπως φαίνεται και στο σχήμα 4.21 στα τετράγωνα γκρι πλαίσια.



Τετραγωνικό Πλέγμα



Εξαγωνικό Πλέγμα

Σχήμα 4.21: Απλή απόδοση τετράφωνου κελιού σε κάθε εντοπισμένη κηλίδα.

Η τεχνική η οποία λύνει σε πολύ μεγάλο ποσοστό αυτό το πρόβλημα είναι η εφαρμογή του Βορονόι διαγράμματος [120]. Το Βορονόι διάγραμμα έχει και στο παρελθόν εφαρμοστεί στο συγκεκριμένο πρόβλημα [114-115].

Ας υποθέσουμε ότι $P = \{p_1, p_2, \dots, p_n\}$ είναι ένα σύνολο σημείων στο επίπεδο. Για να δημιουργηθεί το Βορονόι διάγραμμα, κάθε σημείο του επιπέδου αντιστοιχίζεται με ένα από τα σημεία του P . Πιο συγκεκριμένα, σε κάθε p_i αντιστοιχίζονται όλα τα σημεία του επιπέδου που είναι κοντινότερα σε αυτό από οποιοδήποτε άλλο σημείο του συνόλου P . Για τον σκοπό αυτόν χρησιμοποιείται η ευκλείδεια απόσταση και επομένως με μαθηματικούς όρους του διαγράμματος ορίζεται ως:

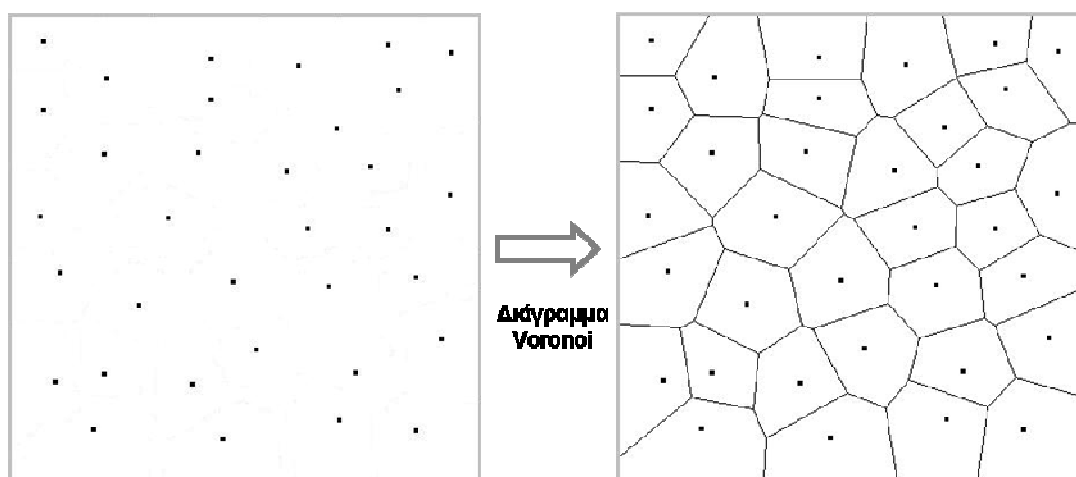
$$V_d(p_i) = \{x : |p_i - x| \leq |p_j - x| \quad \forall j \neq i\}, \quad (4.32)$$

όπου $V_d(p_i)$ είναι το διάγραμμα Βορονόι διάγραμμα ενώ x είναι το σημείο το οποίο θα αντιστοιχιστεί σε ένα από τα σημεία p_i .

Από τα παραπάνω εξάγονται οι βασικές ιδιότητες του Βορονόι διαγράμματος:

1. Το κάθε σημείο p_i περικλείεται από ένα κελί Βορονόι, το οποίο περιέχει σημεία του επιπέδου που είναι πιο κοντά στο p_i .
2. Κάθε ακμή του Βορονόι διαγράμματος αποτελείται από σημεία του επιπέδου τα οποία ισαπέχουν από δύο σημεία του συνόλου P .
3. Κάθε κόμβος του Βορονόι διαγράμματος στον οποίο τέμνονται τρεις ακμές του είναι ένα σημείο του επιπέδου το οποίο ισαπέχει από τρία σημεία του συνόλου P .

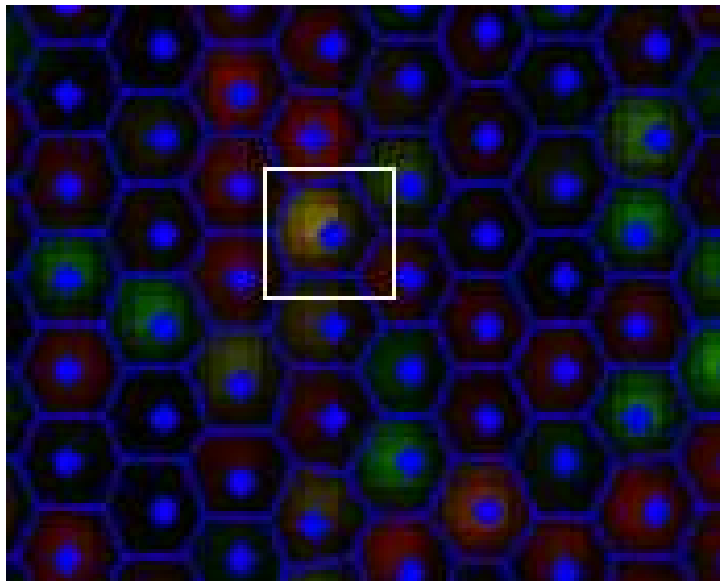
Στο σχήμα 4.22 φαίνεται αριστερά ένα σύνολο σημείων του επιπέδου και δεξιά η εφαρμογή του Βορονόι διαγράμματος για αυτά τα σημεία.



Σχήμα 4.22: Βορονόι διάγραμμα τυχαίων σημείων στο επίπεδο.

Η εφαρμογή του Βορονόι διαγράμματος στις εικόνες μικροσυστοιχιών γίνεται χρησιμοποιώντας τα κέντρα των κηλίδων που έχουν ήδη βρεθεί από τα δύο προηγούμενα βήματα. Κάθε εικονοστοιχείο της εικόνας αντιστοιχίζεται με το πιο κοντινό κέντρο κηλίδας, με αποτέλεσμα τελικά, ένα κελί Βορονόι να δημιουργηθεί γύρω από το κάθε κέντρο. Δεδομένου λοιπόν ότι η κάθε κηλίδα περιέχει τα πιο κοντινά εικονοστοιχεία του κέντρου της κηλίδας, το Βορονόι διάγραμμα αποτελεί μια πολύ αξιόπιστη χωρικά λύση για το πρόβλημα της τοποθέτησης του πλέγματος. Εκτός αυτού η εφαρμογή του Βορονόι διαγράμματος δίνει ακόμα μια λύση, στην περίπτωση όπου τα βήματα εντοπισμού των κηλίδων δεν έχουν εντοπίσει με ακρίβεια το κέντρο. Σε αυτές τις περιπτώσεις όπως είδαμε και παραπάνω το κελί που θα περιέχει την κηλίδα είναι πολύ πιθανό να περιέχει και εικονοστοιχεία από κάποια ή κάποιες γειτονικές κηλίδες. Με την χρήση του Βορονόι

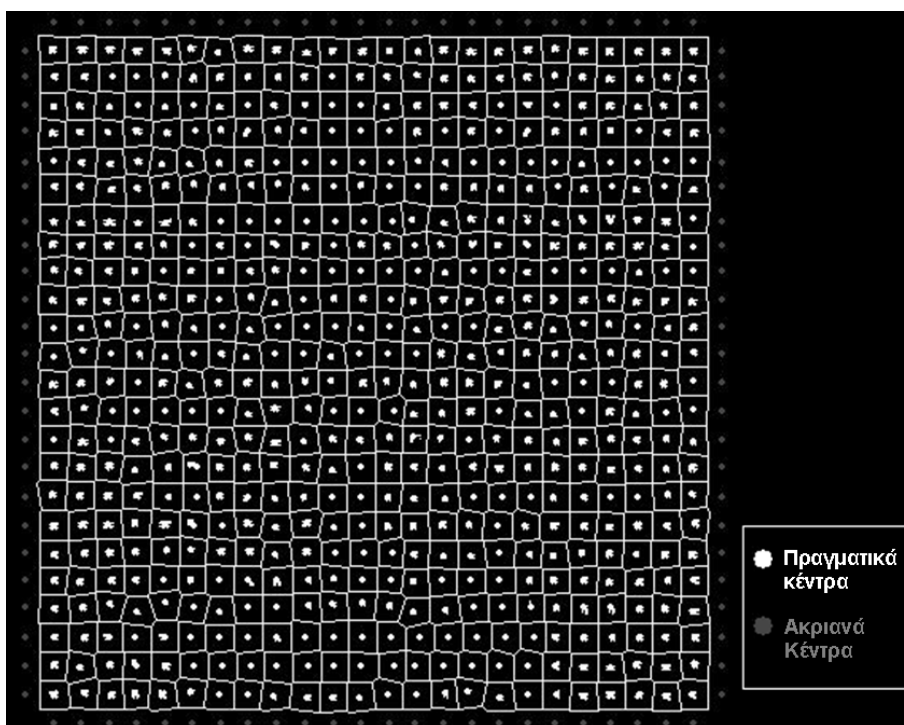
διαγράμματος το κελί της κάθε κηλίδας δεν εξαρτάται μόνο από το κέντρο της, αλλά και από τα κέντρα των γειτόνων της. Έτσι αν μια κηλίδα δεν έχει εντοπιστεί με απόλυτη ακρίβεια, αλλά οι γειτονικές της κηλίδες είναι σωστά εντοπισμένες τότε τα κελιά που θα δημιουργηθούν θα είναι ακριβή. Αυτό φαίνεται στο σχήμα 4.23, στο οποίο το κέντρο της σημειωμένης κηλίδας δεν έχει εντοπιστεί σωστά (βρίσκεται στα όρια της κηλίδας) με την συνεισφορά όμως των γειτονικών κέντρων το Βορονόι διάγραμμα έχει δημιουργήσει ένα κελί που περιέχει εικονοστοιχεία σήματος μόνο από την εν λόγω κηλίδα.



Σχήμα 4.23: Τα κέντρα των γειτονικών κηλίδων συνεισφέρουν σημαντικά ώστε το κελί να τοποθετηθεί σωστά

Ένα από τα βασικά προβλήματα που προκύπτουν από την εφαρμογή του Βορονόι διαγράμματος στις εικόνες των μικροσυστοιχιών είναι η συμπεριφορά του στα όρια της κάθε συστάδας. Τα κέντρα των κηλίδων που βρίσκονται στην πρώτη ή τελευταία στήλη ή γραμμή από κηλίδες της συστάδας, δεν έχουν δίπλα τους άλλες κηλίδες ώστε να περιορίσουν το Βορονόι διάγραμμα. Αυτό έχει ως αποτέλεσμα τα κελιά αυτών των κηλίδων να εκτείνονται μέχρι τα όρια της εικόνας. Για τον λόγο αυτό προσθέσαμε γύρω από την συστάδα τεχνητά κέντρα κηλίδων με την ύπαρξη των οποίων θα περιοριστεί το διάγραμμα. Αυτό φαίνεται στο παρακάτω σχήμα 4.24.

Κάθε κελί Βορονόι θα επεξεργαστεί ένα-ένα ξεχωριστά κατά την διαδικασία της κατάτμησης, ώστε να διαχωριστούν σε αυτό τα εικονοστοιχεία σήματος από τα υποβάθρου και να υπολογιστούν οι ποσοτικοποιημένες τιμές για κάθε μια κηλίδα.



Σχήμα 4.24: Η τοποθέτηση τεχνητών κέντρων γύρω από την συστάδα για τον περιορισμό του Βορονόι διαγράμματος.

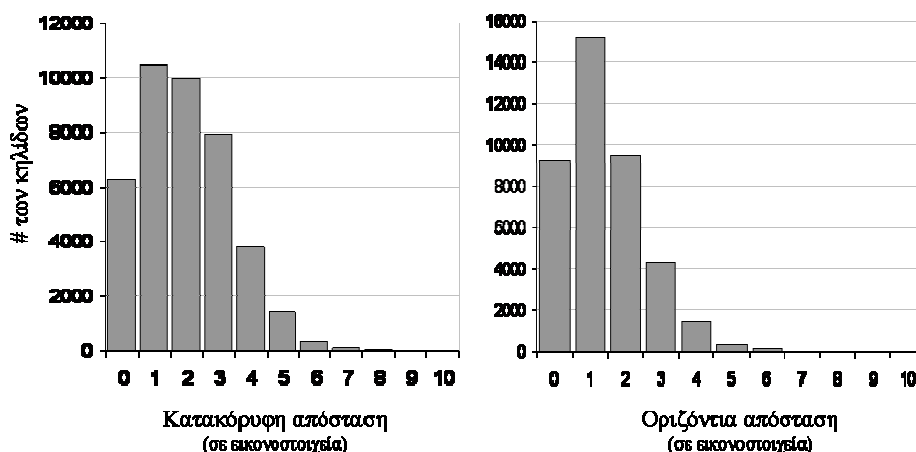
4.6.Αποτελέσματα

Τα αποτελέσματα της παραπάνω μεθοδολογίας θα παρουσιαστούν σε τρία εδάφια. Στο πρώτο εδάφιο παρουσιάζονται τα αποτελέσματα της αρχικής προσέγγισης κηλίδα με κηλίδα που υλοποιήθηκε για εικόνες με τετραγωνικό πλέγμα [179-180]. Στο δεύτερο εδάφιο θα παρουσιαστούν τα αποτελέσματα της μεθοδολογίας σε σύνολο πραγματικών εικόνων με εξαγωνικό πλέγμα [181-182]. Ενώ στο τελευταίο εδάφιο θα παρουσιαστούν αποτελέσματα από τέσσερα διαφορετικά σύνολα εικόνων (και τετραγωνικού και εξαγωνικού πλέγματος) συμπεριλαμβανομένων και προσομοιωμένων εικόνων όπως εξήχθησαν από την γενικευμένη μέθοδο εντοπισμού των κηλίδων [183].

4.6.1. Αρχική προσέγγιση τετραγωνικού πλέγματος

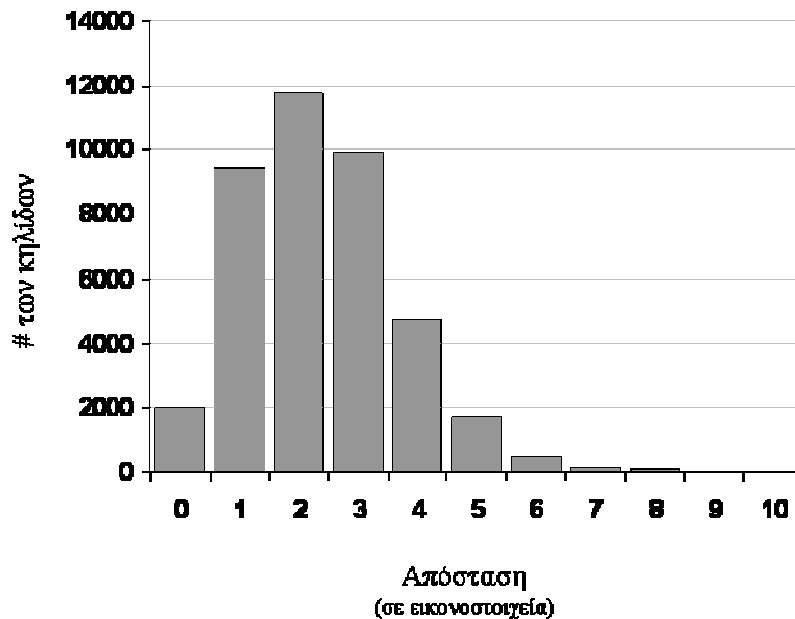
Για την αξιολόγηση της μεθόδου συγκρίναμε τα κέντρα των κηλίδων, τα οποία εντοπίσαμε με την προτεινόμενη μέθοδο, και τα κέντρα από το αρχείο της επισημείωσης της βάσης δεδομένων του Στάνφορντ. Η επισημείωση δεν περιείχε τα κέντρα της κάθε κηλίδας αλλά τις συντεταγμένες που οριοθετούν το τετράγωνο κελί της κάθε κηλίδας. Έτσι εμείς συγκρίναμε τα κέντρα που εντοπίσαμε, με το κέντρο αυτής της τετράγωνης περιοχής το οποίο εύκολα υπολογίσαμε από το αρχείο της επισημείωσης. Πρέπει να επισημάνουμε επίσης ότι στην αρχική μας αυτή προσέγγιση επεξεργαστήκαμε ξεχωριστά τα δύο κανάλια τις εικόνες και έτσι παρουσιάζονται αποτελέσματα και ανά κανάλι.

Στην συνέχεια υπολογίστηκαν οι κατακόρυφες και οι οριζόντιες αποστάσεις (σε εικονοστοιχεία) των κέντρων που υπολογίσαμε με τα κέντρα της επισημείωσης. Από τις αποστάσεις αυτές εξήγαμε τα ιστογράμματα τους, τα οποία μας δίνουν μια πληροφορία για τον αριθμό των κηλίδων που εντοπίστηκαν επιτυχώς και των αριθμό των κηλίδων που εντοπίστηκαν με μικρή απόσταση από τα κέντρα της επισημείωσης. Το ιστογράμμα είναι ένα ραβδόγραμμα των συχνοτήτων εμφάνισης $n_j = \{n_0, n_1, n_2, \dots\}$, για κάθε μία τιμή απόστασης. Για παράδειγμα n_2 είναι ο αριθμός των εντοπισμένων κέντρων τα οποία έχουν βρεθεί 2 εικονοστοιχεία μακριά από το αντίστοιχο επισημειωμένο κέντρο. Τα ιστογράμματα των κατακόρυφων και οριζόντιων αποστάσεων φαίνονται στο σχήμα 4.25



Σχήμα 4.25: Ιστογράμματα των οριζόντιων και κατακόρυφων αποστάσεων των κέντρων των εντοπισμένων με τα κέντρα της επισημείωσης.

Οι οριζόντιες και κατακόρυφες αποστάσεις των κέντρων δεν είναι διαισθητικά αντιληπτές εύκολα. Για τον λόγο αυτό υπολογίσαμε και την δισδιάστατη απόσταση των κέντρων. Πιο συγκεκριμένα υπολογίστηκε η απόσταση Ευκλείδεια (Euclidean) d_i (σε εικονοστοιχεία) των κέντρων που εντοπίστηκαν από την μέθοδο και των επισημειωμένων κέντρων. Το ιστόγραμμα των ευκλείδειων αποστάσεων φαίνεται στο σχήμα 4.26.



Σχήμα 4.26: Ιστόγραμμα των ευκλείδειων αποστάσεων των κέντρων που εντοπίστηκαν με τα επισημειωμένα κέντρα.

Όπως παρατηρούμε και στο σχήμα η συντριπτική πλειοψηφία των αποστάσεων δεν ξεπερνά την απόσταση των πέντε εικονοστοιχείων. Το γεγονός αυτό, σε συνδυασμό με το γεγονός ότι η ακτίνα μιας κηλίδας είναι περίπου 5 εικονοστοιχεία για το συγκεκριμένο σύνολο εικόνων του Στάνφορντ, σημαίνει ότι η συντριπτική πλειοψηφία των κέντρων έχει εντοπιστεί μέσα στα όρια της κηλίδας.

Με σκοπό να εξάγουμε αριθμητικά αποτελέσματα για την αξιολόγηση της μεθόδου υπολογίστηκε η μέση τιμή των αποστάσεων και η τυπική απόκλιση των αποστάσεων d_i , όπως δίνονται από τις παρακάτω σχέσεις:

$$\bar{d} = \frac{1}{N} \sum_{i=1}^N d_i, \quad (4.33)$$

Και

$$\sigma_d = \sqrt{\sum_{i=1}^N \frac{(d_i - \bar{d})^2}{N}}, \quad (4.34)$$

Επίσης με βάση τα παραπάνω υπολογίσαμε την ακρίβεια (Accuracy) της μεθόδου. Θεωρήσαμε δηλαδή ότι η κηλίδες των οποίων τα κέντρα έχουν βρεθεί μέχρι και 5 εικονοστοιχεία μακριά από τα κέντρα της επισημείωσης, έχουν εντοπιστεί σωστά. Η ακρίβεια Acc , της μεθόδου δίνεται από τον τύπο:

$$Acc = \frac{N_{correct}}{N}, \quad (4.35)$$

όπου $N_{correct}$ είναι η αριθμός των σωστά εντοπισμένων κέντρων, ενώ N είναι ο συνολικός αριθμός των κηλίδων της εικόνας.

Πίνακας 4.1: Η ακρίβεια των κέντρων που εντοπιστήκαν με στα όρια της θεωρητικής κηλίδας, η μέση τιμή και η τυπική απόκλιση των αποστάσεων των κέντρων που εντοπίστηκαν με τα κέντρα της επισημείωσης.

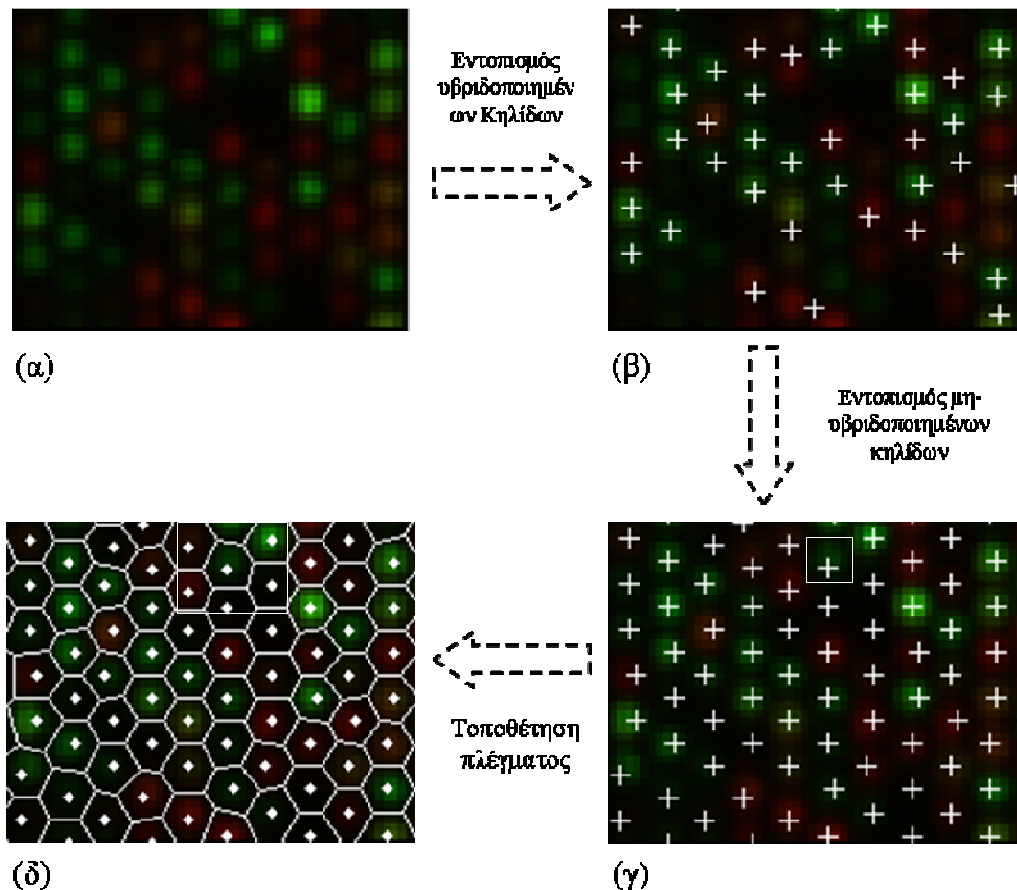
| Κανάλι | Acc (%) | \bar{d} (εικονοστοιχεία) | σ_d (εικονοστοιχεία) |
|---------|--------------|-------------------------------|--------------------------------|
| Κόκκινο | 98.4 | 2.7306 | 1.0985 |
| Πράσινο | 99.8 | 2.3402 | 1.0693 |

Ο πίνακας 4.1 παρουσιάζει τις ολικές τιμές της ακρίβειας, της μέσης τιμής και της τυπικής απόκλισης των αποστάσεων και για τα δύο κανάλια της εικόνας. Όπως παρατηρείται οι ολικές τιμές της ακρίβειας είναι εξαιρετικά υψηλές. Για τις επιμέρους συστάδες οι ακρίβεια κυμάνθηκε από 70,1% έως και 100%. Επίσης μπορούμε να παρατηρήσουμε ότι κατά μέσον όρο τα κέντρα των κηλίδων που εντοπίστηκαν δεν απέχουν πάνω από 3

εικονοστοιχεία απόσταση από τα κέντρα της επισημείωσης, ενώ η τυπική απόκλιση των αποστάσεων αυτών είναι περίπου 1 εικονοστοιχείο.

4.6.2.Εξαγωνικό Πλέγμα

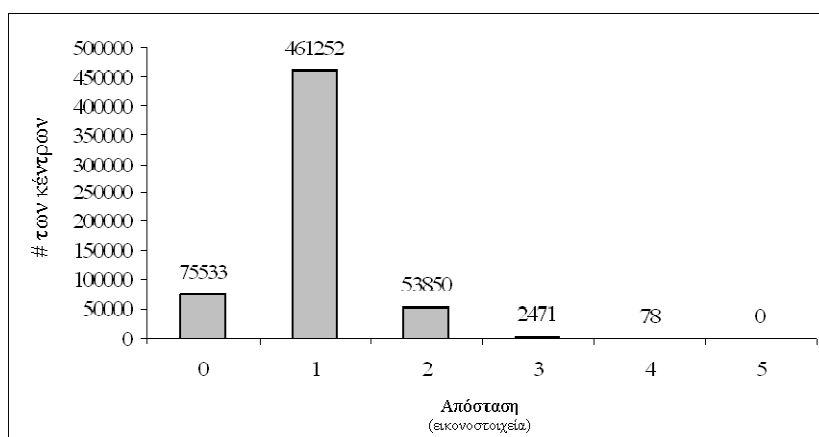
Τα αποτελέσματα που παρουσιάστηκαν στην εργασία [182] αφορούν στην αξιολόγηση της μεθόδου σε σύνολο πραγματικών εικόνων εξαγωνικού πλέγματος πρώτου η μεθοδολογία γενικευτεί ώστε να επεξεργάζεται και εικόνες τετραγωνικού πλέγματος. Για τον σκοπό αυτό χρησιμοποιήθηκαν οι εικόνες από την μικροσυστοιχία CNV370. Κρίνεται σκόπιμο να παρουσιαστούν σε σχήμα τα αποτελέσματα του κάθε βήματος ξεχωριστά σε ένα τμήμα της εικόνας. Τα αποτελέσματα αυτά απεικονίζονται στο σχήμα 4.27.



Σχήμα 4.27: Τα αποτελέσματα της μεθοδολογίας για εικόνες με εξαγωνικό πλέγμα: α) μέρος της αρχικής εικόνας, β) το αποτέλεσμα του εντοπισμού των φωτεινών κηλίδων γ) το αποτέλεσμα του αλγόριθμου των αναπτυσσόμενων ομόκεντρων πολυγώνων δ) το αποτέλεσμα της εφαρμογής του Βορονόι διαγράμματος.

Στο σχήμα 4.27α φαίνεται ένα κομμάτι της αρχικής εικόνας την οποία επεξεργαστήκαμε, ενώ τα σχήματα 4.27β, 4.27γ και 4.27δ δείχνουν τα αποτελέσματα των βημάτων του εντοπισμού των φωτεινών κηλίδων, του εντοπισμού των σκοτεινών κηλίδων και της εφαρμογής του Βορονόι διαγράμματος αντίστοιχα. Στο σχήμα 4.27β οι άσπροι σταυροί αναπαριστούν τα κέντρα μάζας των αντικειμένων που έχουν εντοπιστεί με 8-συνδετικότητα. Όπως παρατηρούμε αυτά είναι κυρίως κέντρα φωτεινών κηλίδων, όπως αυτές απεικονίζονται στην εικόνα. Υπάρχουν βέβαια και φωτεινές κηλίδες για τις οποίες δεν έχει αντιστοιχηθεί κάποιο κέντρο μάζας. Αυτό το φαινόμενο συμβαίνει στις πολύ φωτεινές κηλίδες, οι οποίες κατά το βήμα του εντοπισμού των φωτεινών κηλίδων εντοπίζονται ως ένα αντικείμενο μαζί (συσσωμάτωμα) με τις γειτονικές τους κηλίδες. Από την άλλη μεριά, στο σχήμα 4.27β φαίνονται και σκοτεινές κηλίδες ή σκοτεινές περιοχές στις οποίες θα έπρεπε να υπάρχουν κηλίδες και οι οποίες δεν έχουν εντοπιστεί. Τα κέντρα αυτών των κηλίδων εντοπίζονται στο επόμενο βήμα της μεθόδου όπως φαίνεται στο σχήμα 4.27γ. Τέλος όπως παρουσιάζει το σχήμα 4.27δ η εφαρμογή του Βορονόι διαγράμματος αποδίδει μια περιοχή σε κάθε μια κηλίδα της εικόνας.

Για την αριθμητική αξιολόγηση της μεθόδου, όπως ακριβώς έγινε και για την αρχική μας προσέγγιση σε εικόνες τετραγωνικού πλέγματος, τα κέντρα τα οποία εξήχθησαν από την εφαρμογή της μεθόδου συγκρίθηκαν με αυτά της επισημείωσης του συνόλου εικόνων. Συγκεκριμένα, υπολογίστηκε η ευκλείδεια απόσταση μεταξύ των εντοπισμένων κέντρων και των κέντρων της επισημείωσης, και από αυτές εξήχθησαν τα ιστογράμματα τους.



Σχήμα 4.28: Το ιστόγραμμα των αποστάσεων μεταξύ των εντοπισμένων κέντρων και των κέντρων της επισημείωσης.

Όπως φαίνεται στο σχήμα 4.28 η πλειοψηφία των κέντρων που έχουν εντοπιστεί δεν υπερβαίνουν την απόσταση των δύο εικονοστοιχείων από τα κέντρα της επισημείωσης. Δεδομένου ότι οι εικόνες που χρησιμοποιήσαμε έχουν κηλίδες με ακτίνα 3 εικονοστοιχεία, η πλειοψηφία των κέντρων που εντόπισε η μέθοδος βρίσκεται μέσα στα όρια της κηλίδας.

Σε αντιστοιχία με το τετραγωνικό πλέγμα για να εξάγουμε την ακρίβεια της μεθόδου θεωρήσαμε ότι ένα κέντρο έχει εντοπιστεί σωστά όταν η απόσταση του από το κέντρο της επισημείωσης είναι μικρότερη από την ακτίνα των κηλίδων, δηλαδή από 3 εικονοστοιχεία. Το κατώφλι των 2 εικονοστοιχείων είναι και αυτό βασισμένο όπως και στο τετραγωνικό πλέγμα στο γεγονός ότι τα κέντρα, τα οποία εντοπίστηκαν είναι κέντρα μάζας και όχι χωρικά κέντρα, όπως αυτά της επισημείωσης. Πέραν της ακρίβεια της μεθόδου, όπως έγινε και στην περίπτωση του τετραγωνικού πλέγματος, υπολογίστηκαν η μέση τιμή και η τυπική απόκλιση των αποστάσεων. Η τρεις αυτές τιμές μαζί με τις συχνότητες εμφάνισης των αποστάσεων παρουσιάζονται στον πίνακα 4.2. Η τιμές αυτές εξήχθησαν δεδομένου ότι ο αλγόριθμος εκκινεί από 100 αρχικά σημεία, για τα οποία η μέθοδος παρουσίασε τα καλύτερα αποτελέσματα.

Πίνακας 4.2: Τα αποτελέσματα της προτεινόμενης μεθόδου με χρήση 100 αρχικών σημείων.

| | | | Συχνότητες των αποστάσεων (%) | | | | |
|-------|--------|--------|----------------------------------|-------|-------|-------|-------|
| Acc | Mean | STD | n_0 | n_1 | n_2 | n_3 | n_4 |
| 99.57 | 0.9722 | 0.4850 | 12.73 | 77.76 | 9.08 | 0.42 | 0.01 |

Με σκοπό να διερευνήσουμε τον ρόλο των πολλαπλών αρχικών σημείων του αλγορίθμου πραγματοποιήθηκαν πειράματα με διαφορετικό αριθμό σημείων εκκίνησης. Τα αποτελέσματα της μεθόδου για τα διαφορετικά αυτά πειράματα φαίνονται στον Πίνακα 4.3. Στον πίνακα αυτόν παρουσιάζονται οι τιμές της ακρίβειας της μεθόδου, καθώς επίσης και η μέση τιμή και η τυπική απόκλιση των αποστάσεων των εντοπισμένων κέντρων με τα κέντρα της επισημείωσης, για κάθε περίπτωση διαφορετικού πλήθους σημείων εκκίνησης.

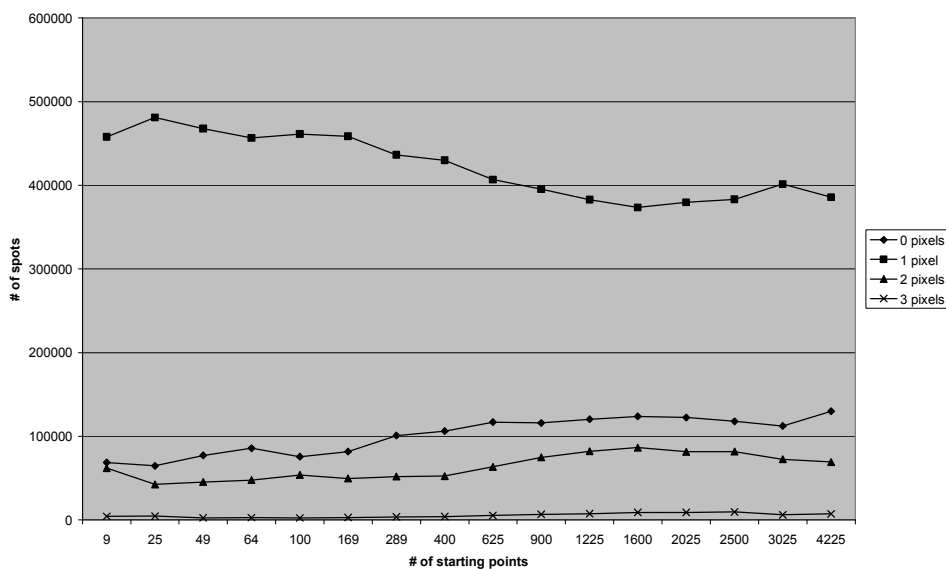
Όπως φαίνεται από τα αποτελέσματα, η προτεινόμενη μέθοδος δίνει πολύ καλά αποτελέσματα, (ακρίβεια περίπου 99%) για εικόνες που χρησιμοποιούν εξαγωνικό πλέγμα για την χωροθέτηση των κηλίδων. Όπως φαίνεται από τους πίνακες 4.2 και 4.3, η πλειονότητα των εντοπισμένων από την μέθοδο κέντρων, είτε έχει με ακρίβεια εντοπιστεί (έχουν απόσταση μηδέν από τα επισημειωμένα κέντρα) είτε βρίσκονται στην 3 επί 3 γειτονιά των επισημειωμένων κέντρων (δηλαδή έχουν απόσταση 1 εικονοστοιχείο). Παρατηρούμε βεβαία ότι τα κέντρα τα οποία έχουν εντοπιστεί με ακρίβεια (μηδενική απόσταση), είναι πολύ λιγότερα σε πλήθος από αυτά τα οποία έχουν βρεθεί σε απόσταση ενός εικονοστοιχείου μακριά. Για παράδειγμα, όπως φαίνεται και στο ιστόγραμμα του σχήματος 4.28, τα κέντρα που είναι με ακρίβεια εντοπισμένα είναι στον αριθμό περίπου 75000, ενώ εκείνα που έχουν απόσταση 1 εικονοστοιχείο από τα επισημειωμένα είναι περίπου 460000.

Πίνακας 4.3: Τα αποτελέσματα της μεθόδου για διαφορετικό πλήθος σημείων εκκίνησης.

| # σημείων εκκίνησης | Συχνότητες αποστάσεων (%) | | | | | | Acc (%) | mean | STD |
|---------------------------|------------------------------|-------|-------|------|------|------|--------------|--------|---------------|
| | n0 | n1 | n2 | n3 | n4 | n5 | | | |
| 9 | 11.60 | 77.19 | 10.46 | 0.75 | 0.01 | 0.00 | 99.24 | 1.0039 | 0.5015 |
| 25 | 10.91 | 81.10 | 7.17 | 0.80 | 0.02 | 0.00 | 99.18 | 0.9792 | 0.4631 |
| 49 | 13.00 | 78.87 | 7.67 | 0.44 | 0.02 | 0.00 | 99.54 | 0.9762 | 0.4739 |
| 64 | 14.46 | 76.96 | 8.07 | 0.49 | 0.02 | 0.00 | 99.49 | 0.9466 | 0.4941 |
| 100 | 12.73 | 77.76 | 9.08 | 0.42 | 0.01 | 0.00 | 99.57 | 0.9722 | 0.4850 |
| 169 | 13.76 | 77.33 | 8.39 | 0.50 | 0.02 | 0.00 | 99.48 | 0.9568 | 0.4912 |
| 289 | 17.01 | 73.55 | 8.76 | 0.65 | 0.03 | 0.00 | 99.32 | 0.9314 | 0.5307 |
| 400 | 17.92 | 72.48 | 8.86 | 0.70 | 0.03 | 0.00 | 99.27 | 0.9244 | 0.5413 |
| 625 | 19.72 | 68.60 | 10.74 | 0.92 | 0.02 | 0.00 | 99.06 | 0.9293 | 0.5820 |
| 900 | 19.55 | 66.65 | 12.60 | 1.15 | 0.04 | 0.00 | 98.80 | 0.9550 | 0.6080 |
| 1225 | 20.29 | 64.55 | 13.84 | 1.28 | 0.04 | 0.00 | 98.68 | 0.9623 | 0.6283 |
| 1600 | 20.86 | 62.95 | 14.58 | 1.54 | 0.07 | 0.00 | 98.39 | 0.9701 | 0.6493 |

| | | | | | | | | | |
|------|-------|-------|-------|------|------|------|-------|---------------|--------|
| 2025 | 20.63 | 63.99 | 13.74 | 1.55 | 0.08 | 0.00 | 98.36 | 0.9648 | 0.6424 |
| 2500 | 19.88 | 64.61 | 13.76 | 1.65 | 0.09 | 0.01 | 98.25 | 0.9749 | 0.6414 |
| 3025 | 18.93 | 67.69 | 12.24 | 1.06 | 0.07 | 0.01 | 98.86 | 0.9568 | 0.6002 |
| 4225 | 21.90 | 65.03 | 11.68 | 1.26 | 0.10 | 0.02 | 98.61 | 0.9270 | 0.6279 |

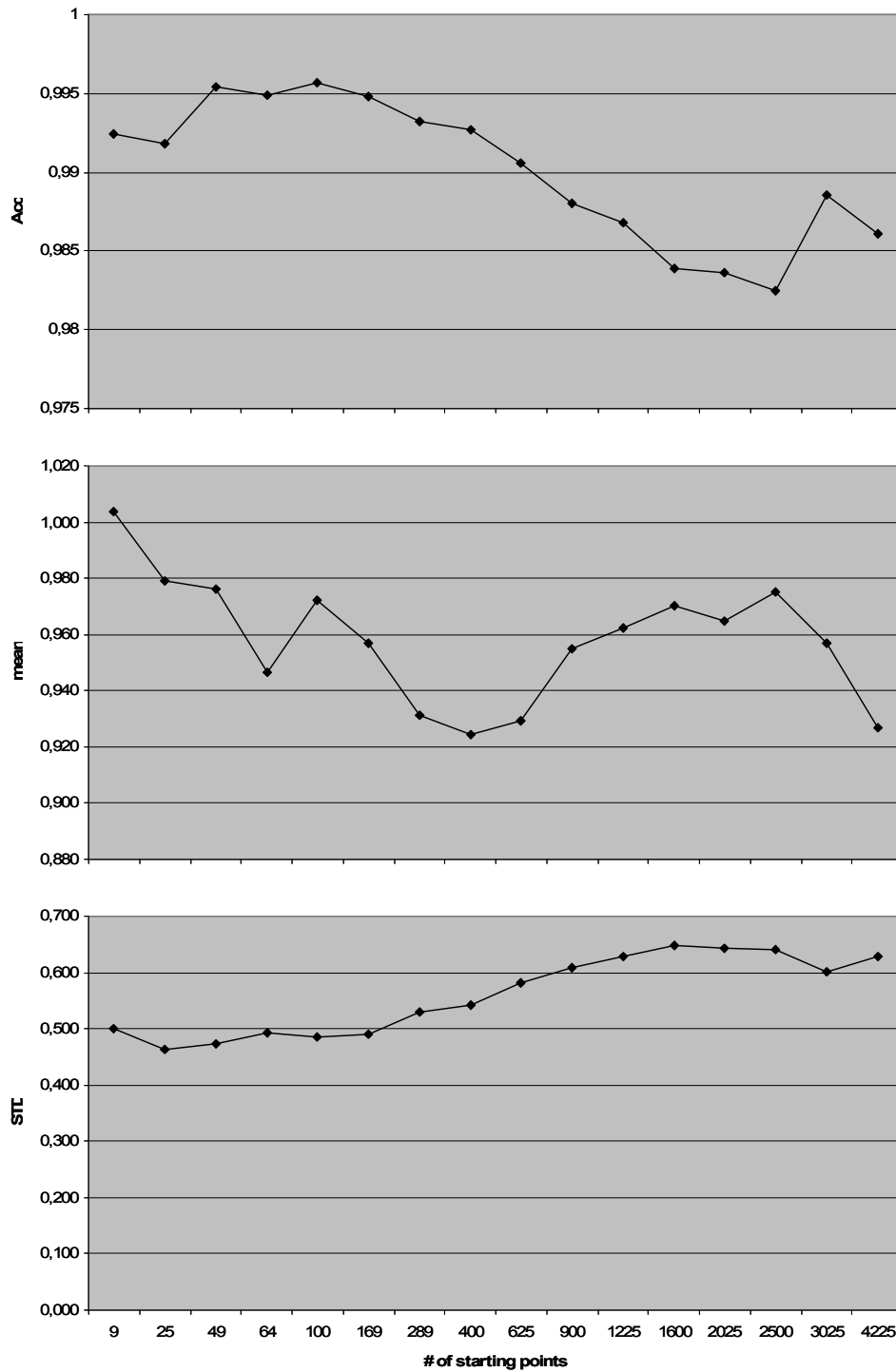
Αυτό το φαινόμενο οφείλεται στο γεγονός ότι η εικόνα είναι ένα ψηφιοποιημένο δισδιάστατο σήμα και η απόσταση ενός εικονοστοιχείου από ένα κεντρικό εικονοστοιχείο μεταφράζεται πρακτικά στο γεγονός ότι το εντοπισμένο κέντρο μπορεί να βρίσκεται σε 8 διαφορετικά εικονοστοιχεία της εικόνας (δηλαδή τα 8 γειτονικά εικονοστοιχεία του κεντρικού εικονοστοιχείου). Αν θεωρηθεί ότι τα κέντρα με απόσταση 1 από τα επισημειωμένα κέντρα είναι ισόποσα καταναμημένα στους 8 γείτονες του κεντρικού εικονοστοιχείου τότε για κάθε έναν από αυτούς αντιστοιχούν περίπου 57500 εικονοστοιχεία, πλήθος το οποίο φαίνεται λογικό δεδομένου ότι τα εντοπισμένα κέντρα με απόλυτη ακρίβεια είναι 75000.



Σχήμα 4.29: Το πλήθος των κέντρων για κάθε απόστασης μεταξύ των εντοπισμένων και επισημειωμένων κέντρων σε συνάρτηση με το πλήθος των σημείων εκκίνησης.

Με σκοπό να απεικονιστούν γραφικά αυτά τα αποτελέσματα, σχεδιάστηκαν τα σχήματα 4.29 και 4.30. Το σχήμα 4.29 είναι το γράφημα του πλήθους των κέντρων που εντοπίστηκαν 0, 1, 2 και 3 εικονοστοιχεία μακριά από τα επισημειωμένα κέντρα, σε

συνάρτηση με το πλήθος των σημείων εκκίνησης. Στο γράφημα αυτό υπάρχουν τέσσερις καμπύλες, μια για κάθε συχνότητα των αποστάσεων. Στο σχήμα 4.30 φαίνεται η μεταβολή της ακρίβειας, της μέσης τιμής και της τυπικής απόκλισης των αποστάσεων σε συνάρτηση με το πλήθος των σημείων εκκίνησης.



Σχήμα 4.30: Η ακρίβεια της μεθόδου, η μέση τιμή και η τυπική απόκλιση των αποστάσεων σε συνάρτηση με το πλήθος των σημείων εκκίνησης.

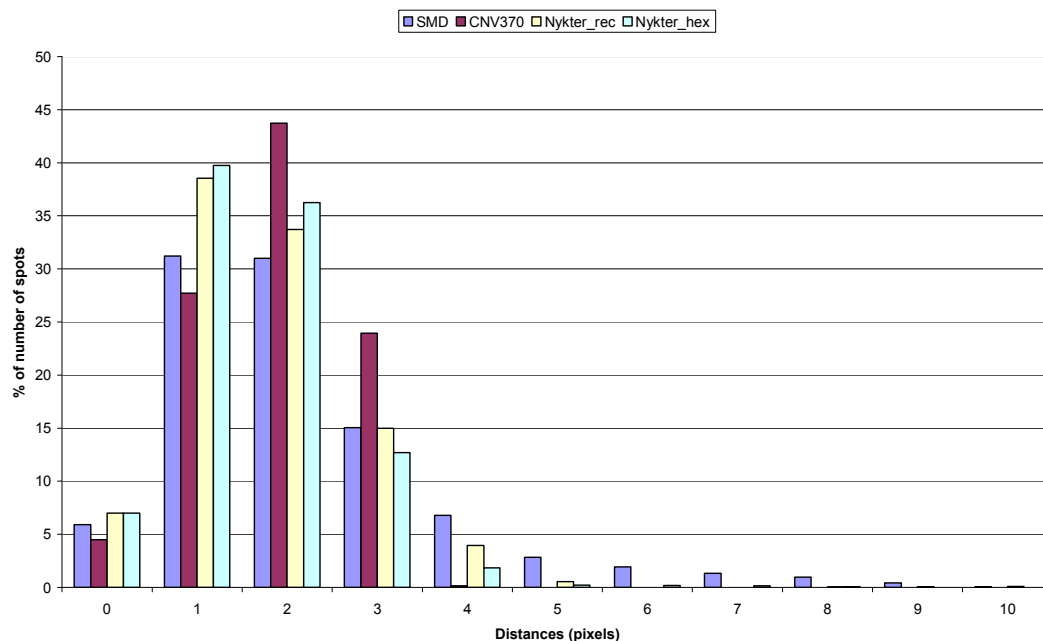
Σχετικά με τον ρόλο των πολλαπλών σημείων εκκίνησης, όπως φαίνεται στο σχήμα 4.30 και στον πίνακα 4.3, η ακρίβεια της μεθόδου αυξάνεται μέχρι ο αριθμός των σημείων εκκίνησης γίνει 100. Όσο μεγαλώνει από εκεί και πέρα το πλήθος των σημείων εκκίνησης η ακρίβεια μειώνεται. Αυτό πιθανότατα οφείλεται στο γεγονός ότι μπορεί να αυξάνεται ο αριθμός των σημείων εκκίνησης που χρησιμοποιεί ο αλγόριθμος και τα οποία είναι λάθος εντοπισμένα από την διαδικασία του εντοπισμού των φωτεινών κηλίδων. Κάποια από αυτά η διαδικασία βελτιστοποίησης αδυνατεί να διορθώσει. Εκτός από την πτώση της ακρίβειας της μεθόδου το φαινόμενο αυτό ευθύνεται για την αύξηση της μέσης τιμής και της τυπικής απόκλισης των αποστάσεων. Σχετικά με τον ρόλο των πολλαπλών σημείων εκκίνησης, άλλο ένα συμπέρασμα εξάγεται από το σχήμα 4.29.

Στο σχήμα αυτό φαίνεται ότι όσο περισσότερα είναι τα σημεία εκκίνησης τόσο αυξάνονται και τα κέντρα τα οποία εντοπίζονται με απόλυτη ακρίβεια (μηδενική απόσταση από τα επισημειωμένα κέντρα). Αντίθετα, τα κέντρα, τα οποία εντοπίζονται στην 3Χ3 γειτονιά των επισημειωμένων κέντρων συνεχώς μειώνονται.

4.6.3.Γενικευμένη Μέθοδος

Στο εδάφιο αυτό θα προσθέσουμε αποτελέσματα τα οποία εξήχθησαν μετά την γενίκευση της μεθόδου για τους δύο τύπους πλέγματος. Για την αξιολόγηση της γενίκευσης χρησιμοποιήθηκαν 4 διαφορετικά σύνολα εικόνων. Συγκεκριμένα χρησιμοποιήθηκαν εικόνες από την βάση δεδομένων του Στάνφορντ, από την μικροσυστοιχία CNV370, καθώς επίσης και προσομοιωμένες εικόνες, οι οποίες παρήχθησαν από τον προσομοιωτή του Nykter [78].

Όπως ακριβώς και στα προηγούμενα αποτελέσματα εξάγαμε τα ιστογράμματα των αποστάσεων μεταξύ των εντοπισμένων και των αντίστοιχων επισημειωμένων κέντρων, για κάθε ένα από τα τέσσερα σύνολα εικόνων. Τα ιστογράμματα παρουσιάζονται στο σχήμα 4.31.



Σχήμα 4.31: Ιστογράμματα των αποστάσεων για κάθε ένα σύνολο εικόνων.

Σε αντιστοιχία με τα προηγούμενα αποτελέσματα υπολογίστηκε για κάθε ένα σύνολο εικόνων η μέση τιμή και η τυπική απόκλιση των αποστάσεων, καθώς επίσης και η ακρίβεια της μεθόδου όπως ορίστηκαν στις εξισώσεις 4.33-4.35 αντίστοιχα. Τα αποτελέσματα παρουσιάζονται στον πίνακα 4.4

Πίνακας 4.4: Αριθμητικά αποτελέσματα τις γενικευμένης μεθοδολογίας σε τέσσερα διαφορετικά σύνολα εικόνων.

| Σύνολο Εικόνων | Περιγραφή | Θεωρητική Ακτίνα | Mean (εικονοστοιχεία) | STD (εικονοστοιχεία) | Acc (%) |
|----------------|------------------------------------|------------------|-----------------------|----------------------|---------|
| SMD | Πραγματικές, τετραγωνικό πλέγμα | 5 | 2.52 | 2.59 | 92.8 |
| Nykter_rec | Προσομοιωμένες, τετραγωνικό πλέγμα | 5 ή 7 | 1.77 | 1.16 | 98.9 |
| CNV370 | Πραγματικές, Εξαγωνικό πλέγμα | 3 | 1.88 | 0.82 | 99.8 |
| Nykter_hex | Προσομοιωμένες, Εξαγωνικό πλέγμα | 5 ή 7 | 1.94 | 2.32 | 97.5 |

Πρέπει να σημειωθεί ότι η ακρίβεια της μεθόδου (*Acc*) υπολογίστηκε για κάθε σύνολο δεδομένων θεωρώντας κάθε φορά ότι το εντοπισμένο κέντρο βρίσκεται εντός των ορίων της θεωρητικής ακτίνας των κηλίδων του εκάστοτε συνόλου. Όπως φαίνεται από τον πίνακα 4.4 η μέθοδος παρουσιάζει υψηλή ακρίβεια κυρίως στα τρία από τα τέσσερα σύνολα εικόνων. Λίγο χαμηλότερα ποσοστά στην ακρίβεια παρουσιάζει η μέθοδος για τις πραγματικές εικόνες τετραγωνικού πλέγματος.

4.7.Συμπεράσματα

Στο παρόν κεφάλαιο μελετήσαμε την μεθοδολογία, η οποία προτείνεται για τον εντοπισμό των κηλίδων και την τοποθέτηση του πλέγματος. Η μεθοδολογία αυτή αποτελείται από τέσσερα βήματα. Αρχικά, εντοπίζονται η συστάδες της εικόνας, οι οποίες απομονώνονται με την χρήση των προβολών της εικόνας στις δύο διευθύνσεις (οριζόντια και κατακόρυφη). Στην συνέχεια σε κάθε μια από τις συστάδες εντοπίζονται τα αντικείμενα υψηλής φωτεινότητας, τα οποία πιθανότατα είναι οι κηλίδες της εικόνας. Το τρίτο βήμα της μεθοδολογίας αφορά στον εντοπισμό των κηλίδων που δεν έχουν εντοπιστεί. Η κηλίδες αυτές είναι σκοτεινές λόγω του χαμηλού βαθμού υβριδοποίησης των ανιχνευτών με τα δείγματα. Ο σκοπός του βήματος αυτού είναι να προσδιοριστούν οι θέσεις στις οποίες βρίσκονται η κηλίδες χαμηλής φωτεινότητας στην εικόνα. Το τελευταίο βήμα της μεθοδολογίας είναι η τοποθέτηση του πλέγματος, δηλαδή η απομόνωση της κάθε κηλίδας μέσα σε ένα κελί. Όπως θα δούμε στο επόμενο κεφάλαιο κάθε κελί από αυτά θα κατατμηθεί ξεχωριστά ώστε να εξαχθούν οι λόγοι υβριδοποίησης των δειγμάτων για κάθε μια κηλίδα της εικόνας. Η μεθοδολογία έχει γενικευτεί ώστε να επεξεργάζεται εικόνες, οι οποίες έχουν είτε τετραγωνική είτε εξαγωνική δομή.

Η διαδικασία, η οποία ακολουθήθηκε για τον εντοπισμό των συστάδων στην εικόνα είναι βασισμένη στην ολιστική προσέγγιση του εντοπισμού των κηλίδων. Το συγκεκριμένο βήμα κάνει χρήση των προβολών της εικόνας, στις οποίες εντοπίζει τις κοιλάδες, με σκοπό να βρεθούν τα σκοτεινά μονοπάτια που διαχωρίζουν τις συστάδες μεταξύ τους. Το βήμα αυτό αρχικά σχεδιάστηκε με σκοπό την επεξεργασία των προβολών ώστε οι κοιλάδες και οι κορυφές τους να γίνουν περισσότερο διακριτές. Έτσι θα ήταν εύκολο να εντοπιστούν από μια αυτόματη μέθοδο διαχωρισμού των συστάδων. Ωστόσο, κατά την γενίκευση της

μεθοδολογίας διαπιστώθηκε ότι η φύση των προβολών μιας εικόνας που κάνει χρήση εξαγωνικού πλέγματος διαφοροποιείται. Η διατριβή προσανατολίστηκε σε μια προσέγγιση, η οποία είναι απαλλαγμένη από κάθε είδους κατώφλι, για να φέρεις εις πέρας τον διαχωρισμό των συστάδων. Ένα σημαντικό σημείο που πρέπει να αναφέρουμε είναι το γεγονός ότι το βήμα του διαχωρισμού των συστάδων όπως ακριβώς υλοποιήθηκε μπορεί να αποτελέσει από μόνο του μια ολιστική μέθοδο για την τοποθέτηση πλέγματος σε εικόνες μικροσυστοιχιών τετραγωνικού πλέγματος. Εφόσον έχουν εντοπιστεί όλες οι κοιλάδες των προβολών (και οι κοιλάδες μικρού όγκου και οι κοιλάδες μεγάλου όγκου) και έχουν διαχωριστεί μεταξύ τους με την χρήση του αλγόριθμου K-μέσων, αρκεί να δημιουργηθούν διαχωριστικές γραμμές στα σημεία που βρίσκονται οι κοιλάδες μικρού όγκου. Αυτό θα γινόταν βέβαια στην περίπτωση που θέλαμε να επεξεργαστούμε μόνο εικόνες τετραγωνικού πλέγματος, όπου οι κηλίδες είναι χωροθετημένες σε γραμμές και στήλες. Επειδή όμως σκοπό μας είναι η δημιουργία μια γενικευμένης μεθοδολογίας και για τους δύο τύπους μικροσυστοιχιών, η επεξεργασία των εικόνων συνεχίστηκε με τον εντοπισμό των θέσεων όλων των κηλίδων (προσέγγιση κηλίδα με κηλίδα).

Για το βήμα του εντοπισμού των φωτεινών κηλίδων μέσα στην κάθε συστάδα υλοποιήθηκαν δύο διαφορετικές μέθοδοι. Όπως αναφέρθηκε και παραπάνω η ακρίβεια σε επίπεδο εικονοστοιχείου αυτού του βήματος δεν είναι το ζητούμενο. Μια απλή μέθοδος κατάτμησης για τον εντοπισμό των φωτεινών κηλίδων αρκεί για να εντοπίσει κηλίδες υψηλής φωτεινότητας στην εικόνα. Η δύο τεχνικές οι οποίες χρησιμοποιήθηκαν είχαν μία βασική διαφορά. Η τεχνική σύμπτωσης προτύπου στόχευε στην εύρεση των αντικειμένων εκείνων, τα οποία πέραν του γεγονότος ότι είναι φωτεινά έχουν και ομοιότητα στο μέγεθος και στο σχήμα με μια τυπική κηλίδα. Με τον τρόπο αυτόν μπορούν να αποκλειστούν τεχνουργήματα τα οποία προέρχονται από θόρυβο. Ωστόσο, εφόσον το μέγεθος των κηλίδων ποικίλει από εικόνα σε εικόνα, αλλά και από κηλίδα σε κηλίδα, η σύγκριση των αντικειμένων με ένα πρότυπο προκαθορισμένου μεγέθους, εκτός του ότι μπορεί να προκαλέσει λάθη χρειάζεται και εκ των προτέρων γνώση. Από την άλλη μεριά, η χρήση της τεχνικής του Ότσου εντοπίζει όλα τα φωτεινά αντικείμενα της εικόνας χωρίς κανέναν περιορισμό (συμπεριλαμβανομένων και των τεχνουργημάτων), έχει όμως το πλεονέκτημα ότι δεν χρησιμοποιεί καμιά εκ των προτέρων γνώση.

Ο εντοπισμός των κηλίδων που δεν έχουν υβριδοποιηθεί είναι μια αναπόφευκτη διαδικασία για τις προσεγγίσεις κηλίδα με κηλίδα. Αυτό συμβαίνει διότι σε μια εικόνα μικροσυστοιχιών πάντα θα υπάρχουν σκοτεινές κηλίδες, οι οποίες έχουν προκύψει από μη υβριδοποιημένα ζεύγη ανιχνευτών-δειγμάτων. Δεδομένου ότι μια προσέγγιση κηλίδα με κηλίδα απαιτεί την ανεύρεση των θέσεων όλων των κηλίδων της εικόνας, οι σκοτεινές κηλίδες πρέπει και αυτές να εντοπιστούν. Για τον σκοπό αυτό κατά την διάρκεια της διδακτορικής διατριβής αναπτύχθηκε αρχικά μια προσέγγιση κηλίδα με κηλίδα για εικόνες τετραγωνικού πλέγματος. Η προσέγγιση αυτή εντοπίζει σε κάθε επανάληψη την θέση μίας κηλίδας λαμβάνοντας υπόψη την θέση γειτονικών φωτεινών κηλίδων. Η ανάγκη μιας πιο γρήγορης προσέγγισης λόγω της αύξησης των ανιχνευτών σε μια μικροσυστοιχία, αλλά και η χρήση του εξαγωνικού πλέγματος οδήγησαν στην δημιουργία ενός έξυπνου αλγορίθμου, ο οποίος εντοπίζει κηλίδες επάνω στο περίγραμμα ομόκεντρων εξαγώνων. Ο αλγόριθμος αυτός τέλος γενικεύτηκε ώστε να επεξεργάζεται και εικόνες με τετραγωνικό πλέγμα. Αυτό έγινε απλώς συμπεριλαμβάνοντας ως δομική μονάδα στον αλγόριθμο και ομόκεντρα τετράγωνα. Έτσι ολοκληρώθηκε ο αλγόριθμος των αναπτυσσόμενων ομόκεντρων πολυγώνων. Το μεγάλο πλεονέκτημα του αλγορίθμου αυτού είναι το γεγονός ότι ξεπερνά τις προσεγγίσεις κηλίδα με κηλίδα, εντοπίζοντας πολλές κηλίδες σε μια επανάληψη, αλλά και το γεγονός ότι μπορεί να εφαρμοστεί και στους δύο τύπους εικόνων με μια απλή αλλαγή της δομικής μονάδας. Το πρώτο πλεονέκτημα δίνει την δυνατότητα της παραλληλοποίησης του αλγορίθμου, στοιχείο πολύ σημαντικό για την επιτάχυνση της επεξεργασίας των εικόνων.

Σημαντικός είναι ο ρόλος της εφαρμογής του Βορονόι διαγράμματος στην τοποθέτηση πλέγματος κυρίως όταν η εικόνα έχει τις κηλίδες χωροθετημένες σε εξαγωνικό πλέγμα. Η εφαρμογή του Βορονόι διαγράμματος πλεονεκτεί για δύο βασικούς λόγους. Η χρήση απλώς μιας τετράγωνης περιοχής προκαθορισμένου μεγέθους για την απομόνωση της κάθε κηλίδας, είναι προβληματική διότι πολλές φορές στις άκρες της τετράγωνης περιοχής εμπεριέχονται εικονοστοιχεία από γειτονικές κηλίδες. Η χρήση του Βορονόι διαγράμματος λύνει το πρόβλημα, αντιστοιχίζοντας στην κάθε κηλίδα μια εξάγωνη περιοχή. Το δεύτερο πλεονέκτημα της χρήσης του Βορονόι διαγράμματος αφορά στο γεγονός ότι έχει την δυνατότητα να αποδίδει την σωστή περιοχή σε μια κηλίδα ακόμα και αν το κέντρο της δεν έχει εντοπιστεί με απόλυτη ακρίβεια, αλλά με μερικά εικονοστοιχεία απόκλιση. Αυτό

συμβαίνει επειδή για την δημιουργία του κελιού που θα περικλείει μια κηλίδα δεν χρησιμοποιούνται μόνο οι συντεταγμένες του κέντρο της ίδιας της κηλίδας αλλά και οι συντεταγμένες των γειτονικών της κηλίδων.

Στα πλαίσια των αποτελεσμάτων της γενικευμένης μεθόδου επιχειρήθηκε μια σύγκριση των προσεγγίσεων που αναπτύχθηκαν στα πλαίσια της διατριβής για τον εντοπισμό των κηλίδων και την τοποθέτηση πλέγματος. Τα αποτελέσματα φαίνονται στον πίνακα 4.5.

Όπως φαίνεται στον πίνακα 4.5 η γενικευμένη μέθοδος παρουσιάζει συγκρίσιμα αποτελέσματα με τις πρώτες προσεγγίσεις. Συγκεκριμένα, η γενικευμένη μέθοδος παρουσιάζει περίπου ίδια ακρίβεια στα τρία από τα τέσσερα σύνολα εικόνων. Όσον αφορά στην σύγκριση της γενικευμένης μεθόδου με την εργασία [182], είναι αναμενόμενα τα αποτελέσματα εφόσον αυτή είναι η μέθοδος η οποία γενικεύτηκε για να επεξεργάζεται και εικόνες τετραγωνικού πλέγματος. Σε σύγκριση με την αρχική προσέγγιση [179-180], παρουσιάζεται διαφοροποίηση στα αποτελέσματα στο πραγματικό σύνολο εικόνων. Το σύνολο εικόνων από την βάση δεδομένων του Στάνφορντ αποτελείται από εικόνες, οι οποίες έχουν παραχθεί τα πρώτα χρόνια της εξέλιξης των μικροσυστοιχιών, με αποτέλεσμα οι εικόνες αυτές να παρουσιάζουν πολύ θόρυβο, αλλά και να έχουν προβλήματα στρέψης.

Η αρχική μας προσέγγιση είχε την δυνατότητα να αντιμετωπίζει την στρέψη των εικόνων, με την χρήση ενός αλγορίθμου, ο οποίος δεν ενσωματώθηκε στην γενικευμένη μεθοδολογία. Όπως αναφέρθηκε στο 3ο κεφάλαιο τα προβλήματα στρέψης των εικόνων των μικροσυστοιχιών έχουν πλέον εξαλειφτεί, και οι νέες μέθοδοι στην βιβλιογραφία δεν περιλαμβάνουν τέτοιες τεχνικές. Τα προβλήματα στρέψης των εικόνων από την βάση δεδομένων του Στάνφορντ είναι και αυτά τα οποία προκαλούν την διαφορά στην ακρίβεια ως 6 περίπου ποσοστιαίες μονάδες.

Πίνακας 4.5: Συγκριτικά αποτελέσματα των προσεγγίσεων που αναπτύχθηκαν στην παρούσα διατριβή για τον εντοπισμό των κηλίδων, με χρήση και των τεσσάρων συνόλων εικόνων.

| Σύνολο εικόνων | Ποσότητα | [179-180] | [182] | Γενικευμένη Μέθοδος ^[183] |
|-------------------|----------|-----------|-------|---|
| SMD | Mean | 1.35 | | 2.52 |
| | Std | 1.30 | - | 2.59 |
| | Acc | 98.7 | | 92.8 |
| Nykter_rec | Mean | 1.74 | | 1.77 |
| | Std | 1.03 | - | 1.17 |
| | Acc | 99.3 | | 99.0 |
| CNV370 | Mean | | 0.10 | 1.88 |
| | Std | - | 0.59 | 0.82 |
| | Acc | | 99.5 | 99.8 |
| Nykter_hex | Mean | | 2.00 | 1.94 |
| | Std | - | 2.30 | 2.32 |
| | Acc | | 97.2 | 97.5 |

Κεφάλαιο 5: Κατάτμηση εικόνων μικροσυστοιχιών

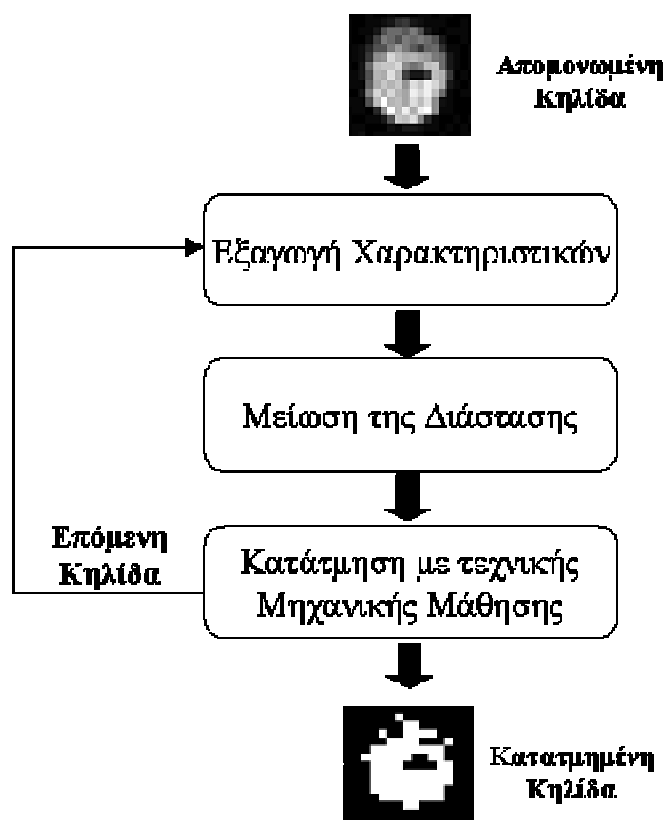
5.Κατάτμηση εικόνων μικροσυστοιχιών

5.1.Εισαγωγή

Στο κεφάλαιο περιγράφεται η μεθοδολογία, η οποία ακολουθήθηκε με σκοπό την κατάτμηση των εικόνων. Η κατάτμηση στοχεύει στον διαχωρισμό των εικονοστοιχείων σήματος, των εικονοστοιχείων υποβάθρου και των εικονοστοιχείων που ανήκουν σε κάποιο τεχνούργημα της εικόνας [81,185]. Το στάδιο αυτό εφαρμόζεται σε κάθε μια κηλίδα της εικόνας ξεχωριστά, καθώς επεξεργάζονται τα κελία που έχουν προκύψει από το προηγούμενο στάδιο. Η κατάτμηση που προτείνεται είναι βασισμένη στην εφαρμογή τεχνικών Μηχανικής Μάθησης [186]. Ήδη όπως αναφέρθηκε και στην βιβλιογραφική ανασκόπηση, οι τεχνικές ομαδοποίησης για το συγκεκριμένο πρόβλημα έχουν επικρατήσει τα τελευταία χρόνια, δεδομένου ότι η κατάτμησης μιας εικόνας εμπίπτει στα πλαίσια της ομαδοποίησης προτύπων. Η συνεισφορά της παρούσας διδακτορικής διατριβής συνίσταται κυρίως σε τρεις τομείς, ι) στα χαρακτηριστικά που χρησιμοποιήθηκαν και τα οποία θα τροφοδοτήσουν της τεχνικές μηχανικής μάθησης, ιι) στο γεγονός ότι για πρώτη φορά εφαρμόστηκαν τεχνικές ταξινόμησης για την κατάτμηση της εικόνας με επίβλεψη, και ιιι) στο γεγονός ότι για πρώτη φορά χρησιμοποιήθηκε τρίτη κλάση εικονοστοιχείων στην

οποία και δόθηκε φυσική υπόσταση. Είναι η κλάση η οποία περιέχει εικονοστοιχεία τα οποία δεν πρέπει να ληφθούν υπόψη κατά την ποσοτικοποίηση της υβριδοποίησης.

Παρακάτω περιγράφονται αναλυτικά τα τρία βασικά βήματα όλης της διαδικασίας της κατάτμηση που ακολουθήθηκαν, την εξαγωγή των χαρακτηριστικών, την μείωση της διάστασης του διανύσματος χαρακτηριστικών, και τέλος την κατάτμηση της εικόνας με τον χαρακτηρισμό των εικονοστοιχείων. Τα βήματα αυτά φαίνονται στο διάγραμμα ροής της μεθοδολογίας το οποίο απεικονίζεται στο σχήμα 5.1.



Σχήμα 5.1: Διάγραμμα ροής της μεθοδολογίας για την κατάτμηση.

5.2. Εξαγωγή των χαρακτηριστικών

Το πρώτο βήμα της κατάτμησης σχετίζεται με την εξαγωγή των χαρακτηριστικών των εικονοστοιχείων, τα οποία θα τροφοδοτήσουν είτε τους αλγόριθμους ομαδοποίησης είτε τους ταξινομητές. Πληθώρα μεθόδων [154-155, 157, 169] έχουν υλοποιηθεί για την κατάτμηση των κηλίδων, οι οποίες κάνουν χρήση αλγορίθμων ομαδοποίησης, όπως όμως

αναφέρθηκε στο 3^ο κεφάλαιο οι μέθοδοι αυτοί τροφοδοτούν τους αλγορίθμους μόνο με την φωτεινότητα των εικονοστοιχείων. Η προτεινόμενη μεθοδολογία δημιούργησε ένα διάλυμα χαρακτηριστικών για κάθε εικονοστοιχείο, στοχεύοντας στην πιο αξιόπιστη κατάτμηση τους. Τα χαρακτηριστικά τα οποία χρησιμοποιήθηκαν είναι στο σύνολο τους 14 και μπορούν να χωριστούν σε τρεις κύριες κατηγορίες [187-189]. Η πρώτη κατηγορία σχετίζεται με τις τιμές της φωτεινότητας του εικονοστοιχείου και των γειτόνων του. Η δεύτερη κατηγορία με την θέση του εικονοστοιχείου μέσα στο κελί της κάθε κηλίδας. Ενώ τέλος η κατανομή και το σχήμα μιας κηλίδας όπως αυτή προσεγγίζεται θεωρητικά χαρακτηρίζει την τρίτη κατηγορία χαρακτηριστικών. Τις κατηγορίες αυτές χαρακτηριστικών τις έχουμε ονομάσει χαρακτηριστικά φωτεινότητας, χωρικά χαρακτηριστικά, και χαρακτηριστικά σχήματος αντίστοιχα.

Χαρακτηριστικά Φωτεινότητας: Η φωτεινότητα του εικονοστοιχείου είναι το χαρακτηριστικό το οποίο κατεξοχήν χρησιμοποιείται για τις μεθόδους κατάτμησης οι που κάνουν χρήση τεχνικών Μηχανικής Μάθησης. Όσο περισσότερο φωτεινό είναι ένα εικονοστοιχείο τόσο αυξάνεται η πιθανότητα το εικονοστοιχείο αυτό να είναι εικονοστοιχείο σήματος. Πέραν τούτου όμως πολύ σημαντική πληροφορία μπορούν να προσφέρουν και η τιμές της φωτεινότητας της κοντινής γειτονιάς του συγκεκριμένου εικονοστοιχείου. Μεγέθη όπως η μέση τιμή της, η τυπική απόκλιση και η εντροπία της γειτονιάς του εικονοστοιχείου μπορούν να προσφέρουν σημαντική πληροφορία. Για παράδειγμα όταν η μέση τιμή της γειτονιάς ενός εικονοστοιχείου είναι υψηλή υπάρχουν πολλές πιθανότητες το εικονοστοιχείο αυτό να είναι εικονοστοιχείο σήματος διότι φαίνεται ότι βρίσκεται σε ένα φωτεινό αντικείμενο της εικόνας. Αντίθετα, αν για ένα εικονοστοιχείο, που το ίδιο έχει υψηλή φωτεινότητα, η μέση τιμή της γειτονιάς του είναι χαμηλή, είναι πιθανό η φωτεινότητα του εικονοστοιχείου να είναι υψηλή λόγω κρουστικού θορύβου (salt-and-pepper noise). Ένα τέτοιο εικονοστοιχείο μπορεί να σημαίνει επίσης ότι βρίσκεται στα όρια της κηλίδας. Αντίστοιχα η μικρή τυπική απόκλιση και η μικρή εντροπία της γειτονιάς αυξάνουν την πιθανότητα το εικονοστοιχείο να βρίσκεται μέσα σε κάποια κηλίδα. Στην παρούσα εργασία μόνο η 3X3 γειτονιά των εικονοστοιχείων χρησιμοποιήθηκε, ενώ θα μπορούσε ενδεχομένως να χρησιμοποιηθεί και η 5X5 γειτονιά. Το θεωρητικό όμως μέγεθος μιας κηλίδας (το οποίο συνήθως δεν ξεπερνά τα 10 εικονοστοιχεία σε διάμετρο) δεν αφήνει περιθώρια για την χρησιμοποίηση μεγαλύτερων

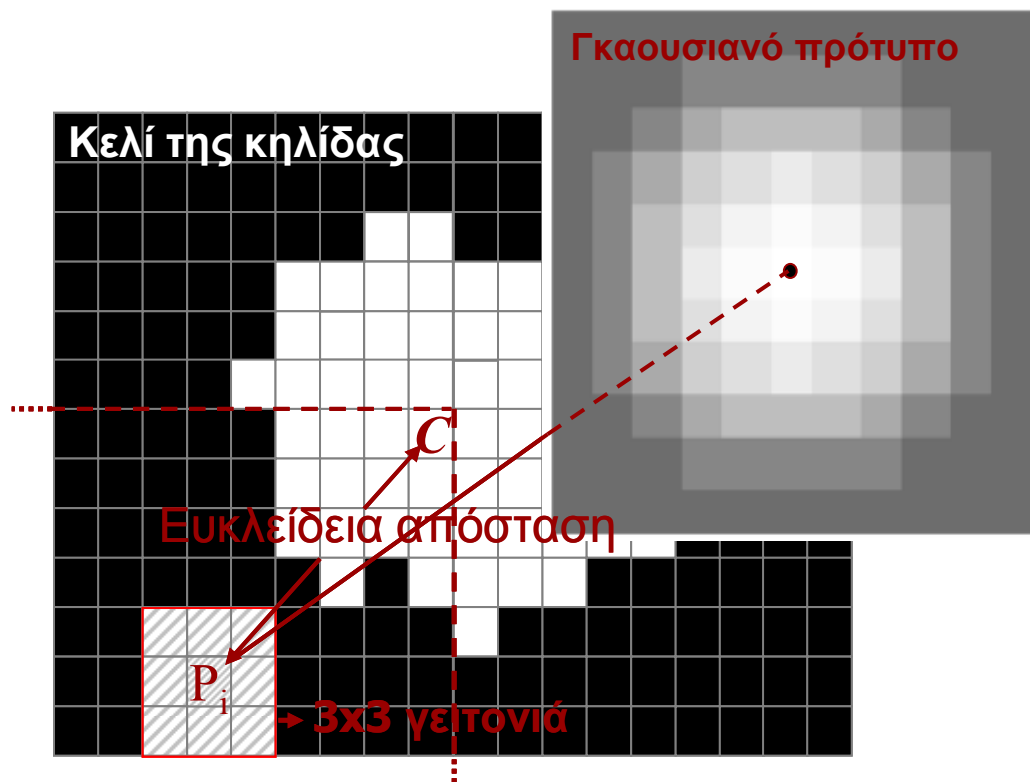
γειτονιών από εικονοστοιχεία. Η χρήση και άλλων χαρακτηριστικών, εκτός από αυτά που σχετίζονται με την φωτεινότητα του εικονοστοιχείου, γίνεται απαραίτητη εφόσον θέλουμε να αποκλείσουμε εικονοστοιχεία που προέρχονται από τεχνουργήματα. Τα εικονοστοιχεία αυτά είναι εξίσου φωτεινά με τα εικονοστοιχεία που ανήκουν σε κηλίδες, ενώ τίποτα δεν περιορίζει ένα τεχνουργήμα να είναι παρόμοιου μεγέθους με μια κηλίδα.

Χωρικά χαρακτηριστικά: Τα χαρακτηριστικά αυτά περιλαμβάνουν τις συντεταγμένες του εικονοστοιχείου στο κελί της κηλίδας. Δύο εικονοστοιχεία με κοντινές συντεταγμένες είναι πιθανό να ανήκουν στην ίδια κατηγορία. Εκτός των συντεταγμένων ως χωρικό χαρακτηριστικό ορίζεται και η απόσταση του εικονοστοιχείου από το κέντρο της κηλίδας. Το κέντρο της κηλίδας είναι γνωστό από την διαδικασία της εύρεσης των κηλίδων. Μικρή απόσταση από το κέντρο της κηλίδας αποτελεί ένδειξη ότι το εικονοστοιχείο ανήκει στα εικονοστοιχεία σήματος, ενώ αντίθετα μεγάλη απόσταση μπορεί να σημαίνει ότι το εικονοστοιχείο είναι εικονοστοιχείο υποβάθρου. Κατ' αυτόν τον τρόπο, η απόσταση του εικονοστοιχείου από το κέντρο της κηλίδας είναι εκείνο το χαρακτηριστικό, το οποίο συνεισφέρει περισσότερο από τα άλλα στον διαχωρισμό των εικονοστοιχείων σήματος με τα εικονοστοιχεία που ανήκουν σε κάποιο τεχνουργήμα. Συνήθως τα τεχνουργήματα παρόλο που είναι πολύ φωτεινά, δεν βρίσκονται στο κέντρο του κελιού της κηλίδας αλλά κοντά στα όρια του.

Χαρακτηριστικά σχήματος: Αυτή η κατηγορία περιέχει ένα χαρακτηριστικό, το οποίο υπολογίζει την ομοιότητα της γειτονιάς του εικονοστοιχείου με την κατανομή μιας θεωρητικής κηλίδας. Το σχήμα και η κατανομή των φωτεινοτήτων των εικονοστοιχείων μέσα σε μια θεωρητική κηλίδα μπορεί με αρκετά καλό τρόπο να προσεγγιστεί από μια δυσδιάστατη Γκαουσιανή κατανομή. Επίσης το μέγεθος μιας κηλίδας είναι συνήθως περίπου 10 εικονοστοιχεία διάμετρος. Με δεδομένα τα παραπάνω δημιουργήθηκε μια δυσδιάστατη Γκαουσιανή σχηματομορφή μεγέθους 11X11 της οποίας η ομοιότητα εξετάζεται με την 11X11 γειτονιά του εικονοστοιχείου. Ο βαθμό ομοιότητας τους εκφράζεται από τον συντελεστή συσχέτισης (correlation coefficient) μεταξύ των δύο. Ο συντελεστής συσχέτισης παίρνει τιμές από -1 έως 1 και όσο προσεγγίζει την μονάδα τόσο μεγαλύτερη είναι οι ομοιότητα των δύο αυτών περιοχών. Έτσι, όσο μεγαλύτερος είναι ο

συντελεστής συσχέτισης τόσο περισσότερο πιθανό είναι τα εικονοστοιχείο να βρίσκεται μέσα σε μια κηλίδα και αντιστρόφως.

Όλα τα παραπάνω χαρακτηριστικά εκτός από τα χαρακτηριστικά θέσης λαμβάνονται και από τα δύο κανάλια της εικόνας. Αντιθέτως, τα χαρακτηριστικά θέσης είναι τα ίδια και για τα δύο κανάλια. Συνοπτικά τα χαρακτηριστικά των εικονοστοιχείων αναγράφονται στον παρακάτω πίνακα 5.1, στον οποίον υπάρχουν και όλοι οι μαθηματικοί τους τύποι. Επίσης το σχήμα 5.2 απεικονίζει ένα κελί μιας κηλίδας και τον τρόπο με τον οποίο εξάγονται τα χαρακτηριστικά.



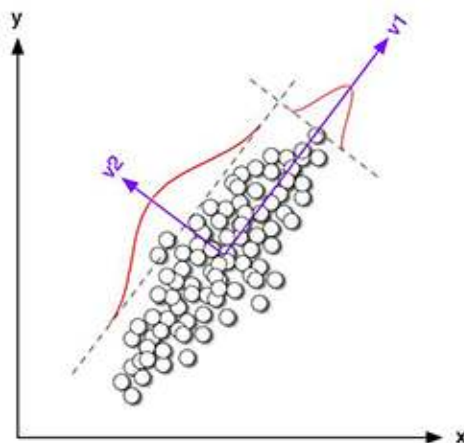
Σχήμα 5.2: Η διαδικασία εξαγωγής των χαρακτηριστικών. Τα χαρακτηριστικά φωτεινότητας σχετίζονται με την φωτεινότητα του P_i και της 3x3 γειτονιάς του και για τα δύο κανάλια της εικόνας. Τα χαρακτηριστικά θέσης είναι οι συντεταγμένες του εικονοστοιχείου P_i και οι αποστάσεις του εικονοστοιχείου από το κέντρο C. Τα χαρακτηριστικά σχήματος είναι ο συντελεστής συσχέτισης της 11X11 διδιάστατης κανονικής κατανομής με την 11X11 γειτονιά του P_i και για τα δύο κανάλια της εικόνας.

Πίνακας 5.1: Τα χαρακτηριστικά που εξάγονται για κάθε εικονοστοιχείο της εικόνας

| Κατηγορία Χαρακτ. | Κανάλι | Περιγραφή | Τύπος |
|-------------------|----------------------|--|--|
| 1 | | Φωτεινότητα του εικονοστοιχείου (i,j) στο κόκκινο κανάλι | $I_r(i,j)$ |
| 2 | | Μέση τιμή της φωτεινότητας του 3-επί-3 κεντροθετημένου παραθύρου | $\bar{I}_r^{Window} = \frac{1}{9} \sum_{i=-1}^1 \sum_{j=-1}^1 I_r(i,j)$ |
| 3 | Κόκκινο | Τυπική απόκλιση της φωτεινότητας του 3-επί-3 κεντροθετημένου παραθύρου | $\sigma_r = \sqrt{\frac{1}{9} \sum_{i=-1}^1 \sum_{j=-1}^1 (I_r(i,j) - \bar{I}_r^{Window})^2}$ |
| 4 | | Εντροπία του 3-επί-3 κεντροθετημένου παραθύρου | $E_r = \sum_{i=-1}^{+1} \sum_{j=-1}^{+1} p_{r,i,j} \log_2 \left(\frac{1}{p_{r,i,j}} \right)$ |
| 5 | Χαρακτ. Φωτεινότητας | Φωτεινότητα του εικονοστοιχείου (i,j) στο πράσινο κανάλι | $I_g(i,j)$ |
| 6 | | Mean intensity value of the 3x3 pixel centered window | $\bar{I}_g^{Window} = \frac{1}{9} \sum_{i=-1}^1 \sum_{j=-1}^1 I_g(i,j)$ |
| 7 | Πράσινο | Τυπική απόκλιση της φωτεινότητας του 3-επί-3 κεντροθετημένου παραθύρου | $\sigma_g = \sqrt{\frac{1}{9} \sum_{i=-1}^1 \sum_{j=-1}^1 (I_g(i,j) - \bar{I}_g^{Window})^2}$ |
| 8 | | Εντροπία του 3-επί-3 κεντροθετημένου παραθύρου | $E_g = \sum_{i=-1}^{+1} \sum_{j=-1}^{+1} p_{g,i,j} \log_2 \left(\frac{1}{p_{g,i,j}} \right)$ |
| 9 | Κόκκινο | Συντελεστής συσχέτισης μεταξύ του 11-επί-11 κεντροθετημένου παραθύρου και του Γκαουσιανού προτύπου I | $r_i = \frac{\sum_{i=-5}^5 \sum_{j=-5}^5 (I_r(i,j) - \bar{I}_r^{Window}) (T(i,j) - \bar{T})}{\sqrt{(\sum_{i=-5}^5 \sum_{j=-5}^5 (I_r(i,j) - \bar{I}_r^{Window})^2) (\sum_{i=-5}^5 \sum_{j=-5}^5 (T(i,j) - \bar{T})^2)}}$ |
| 10 | Πράσινο | Συντελεστής συσχέτισης μεταξύ του 11-επί-11 κεντροθετημένου παραθύρου και του Γκαουσιανού προτύπου I | $r_j = \frac{\sum_{i=-5}^5 \sum_{j=-5}^5 (I_g(i,j) - \bar{I}_g^{Window}) (T(i,j) - \bar{T})}{\sqrt{(\sum_{i=-5}^5 \sum_{j=-5}^5 (I_g(i,j) - \bar{I}_g^{Window})^2) (\sum_{i=-5}^5 \sum_{j=-5}^5 (T(i,j) - \bar{T})^2)}}$ |
| 11 | | Η σειρά στην οποία αφήνει το εικονοστοιχείο στην εικόνα | i |
| 12 | | Η στήλη στην οποία αφήνει το εικονοστοιχείο στην εικόνα | j |
| 13 | Κόκκινο/ Πράσινο | Ενκλειδία απόσταση μεταξύ του εικονοστοιχείου και του κέντρου της κηλίδας | $D_{row} = \sqrt{(i-i_c)^2 + (j-j_c)^2}$, i_c : Row of the spot center, j_c : Column of the spot center. |
| 14 | | Μισγώσιμη απόσταση μεταξύ του εικονοστοιχείου και του κέντρου της κηλίδας | $D_{ij} = i-i_c + j-j_c $ |

5.3.Μείωση της διάστασης των χαρακτηριστικών με χρήση Ανάλυσης Κύριων Συνιστωσών.

Δεδομένου ότι μια εικόνα μικροσυστοιχιών είναι κατεξοχήν μια μεγάλη σε μέγεθος εικόνα η διαδικασία της ταξινόμησης θα πρέπει να επιταχυνθεί μειώνοντας την διάσταση του διανύσματος χαρακτηριστικών. Το γεγονός ότι κάποια από τα χαρακτηριστικά που χρησιμοποιούμε δεν είναι ανεξάρτητα μεταξύ τους, κάνει την χρήση της Ανάλυσης Κύριων Συνιστωσών (Principal Component Analysis – PCA) [190] την η καλύτερη επιλογή για να επιτύχουμε το σκοπό αυτό. Η Ανάλυση Κύριων Συνιστωσών προβάλλει τα χαρακτηριστικά που έχουν εξαχθεί σε ένα σύστημα αξόνων, οι κατευθύνσεις του οποίου μεγιστοποιούν την διασπορά. Οι κατευθύνσεις αυτές προσδιορίζονται από τα ιδιοδιάνυσμα του πίνακα συνμεταβλητότητας, τα οποία αντιστοιχούν στις μεγαλύτερες ιδιοτιμές. Καταυτόν τον τρόπο τα τελικά χαρακτηριστικά τα οποία θα τροφοδοτήσουν τις τεχνικές μηχανικής μάθησης είναι ένας γραμμικός συνδυασμός των αρχικών χαρακτηριστικών. Συγκεκριμένα η ανάλυση κύριων συνιστωσών μειώνει την πλεονάζουσα πληροφορία που περιέχεται στα δεδομένα δημιουργώντας νέα σειρά προτύπων στην οποία οι άξονες του νέου συστήματος συντεταγμένων τείνουν προς την κατεύθυνση που μειώνεται η διακύμανση. Η μέση τιμή των αρχικών δεδομένων ορίζει και την αρχή του μετασχηματισμένου συστήματος, οι άξονες του οποίου είναι αμοιβαία ορθογώνιοι. Στο σχήμα 5.3 φαίνεται ένα παράδειγμα όπου ο δισδιάστατος χώρος μετασχηματίζεται από την ανάλυση κυρίων συνιστωσών με σκοπό οι νέοι άξονες να ελαχιστοποιούν την διακύμανση των δεδομένων. Το μετασχηματισμένο σύστημα αξόνων είναι το σύστημα με άξονες τους V_1 και V_2 .



Σχήμα 5.3: Δισδιάστατο παράδειγμα κατά το οποίο τα οι άξονες μετασχηματίζεται ελαχιστοποιώντας την διακύμανση

Στο πρώτο στάδιο της υλοποίηση του αλγορίθμου υπολογίζεται ο πίνακας συνδιακύμανσης C , των αρχικών δεδομένων. Από τον πίνακα συνδιακύμανσης, προσδιορίζονται οι ιδιοτιμές l_i σύμφωνα με την εξίσωση:

$$|C - l_i I| = 0, \quad i \in [1, 2, K, n] \quad (5.1)$$

όπου n είναι το πλήθος των αρχικών προτύπων και I είναι ο μοναδιαίος πίνακας. Τα ιδιοδιανύσματα e_i ορίζουν τους άξονες των συνιστωσών και προκύπτουν από την εξίσωση:

$$(C - l_i I)e_i = 0. \quad (5.2)$$

Οι κύριες συνιστώσες παράγονται από τη σχέση:

$$P = T D, \quad (5.3)$$

όπου D είναι ο πίνακας των αρχικών δεδομένων και T είναι ο πίνακας μετασχηματισμού που δίνεται από τον τύπο:

$$T = \begin{pmatrix} e_{11} & L & e_{1n} \\ \vdots & & \vdots \\ e_{k1} & M & O \\ \vdots & & \vdots \\ e_{n1} & L & e_{nn} \end{pmatrix} \quad (5.4)$$

Η γραμμικοί συνδυασμοί των χαρακτηριστικών που προκύπτουν από την ανάλυση κύριων συνιστωσών είναι ανεξάρτητοι μεταξύ τους και ταξινομημένοι κατά φθίνουσα διακύμανση. Ο πίνακας διασποράς των μετασχηματισμένων δεδομένων είναι διαγώνιος με στοιχεία που συνίστανται από τις ιδιοτιμές. Η ποσοστιαία συνολική διακύμανση είναι ένα μέτρο με το οποίο καθορίζεται πολλές φορές πόσους από τους γραμμικούς συνδυασμούς των δεδομένων θα λάβουμε τελικά. Αυτή δίνεται από την σχέση:

$$\text{Var} = \frac{l_i}{\sum_{k=1}^n l_k} 100\%. \quad (5.5)$$

Από τα 14 χαρακτηριστικά που αρχικά χρησιμοποιήθηκαν και για ποσοστό $\text{Var}=95\%$, τελικά τις τεχνικές μηχανικής μάθησης που χρησιμοποιήσαμε τις τροφοδότησαν 5 ή 6 γραμμικοί συνδυασμοί τους.

5.4. Διαχωρισμός Εικονοστοιχείων με χρήση τεχνικών Μηχανικής Μάθησης

Η κατάτμηση μιας εικόνας εξ ορισμού είναι η διαδικασία κατά την οποία εικονοστοιχεία με παρόμοια χαρακτηριστικά ομαδοποιούνται. Κατά συνέπεια, εύλογα το συγκεκριμένο πρόβλημα προσεγγίζεται τα τελευταία χρόνια με τεχνικές ομαδοποίησης, οι οποίες εφαρμόζονται στα εικονοστοιχεία της εικόνας. Ένα βασικό μειονέκτημα των εξελιγμένων τεχνικών επεξεργασίας εικόνας είναι το γεγονός ότι χρησιμοποιούν κυρίως χαρακτηριστικά φωτεινότητας των εικονοστοιχείων και είναι πολύ δύσκολο έως αδύνατο να ενσωματωθούν άλλα χαρακτηριστικά που προκύπτουν από την φύση του εκάστοτε προβλήματος. Με την χρήση τεχνικών ομαδοποίησης νέα χαρακτηριστικά μπορούν να ενσωματωθούν στο εξαγόμενο διάνυσμα χαρακτηριστικών όπως είδαμε στο δεύτερο εδάφιο αυτού του κεφαλαίου. Ένα άλλο πλεονέκτημα της χρήσης των τεχνικών ομαδοποίησης σε σχέση με τις τεχνικές επεξεργασίας εικόνας αφορά στον αριθμό των ομάδων που ομαδοποιείς τα εικονοστοιχεία. Ανάλογα με το πρόβλημα και την φύση των εικόνων το ζητούμενο της κατάτμησης μπορεί να είναι είτε ο διαχωρισμός των εικονοστοιχείων σε έναν προκαθορισμένο αριθμό ομάδων, είτε η μελέτη όλων των ομάδων που θα προκύψουν ανεξαρτήτως του αριθμού. Τέτοιου είδους προβλήματα ευνοούνται από την χρήση των τεχνικών ομαδοποίησης για την κατάτμηση, είτε θέτοντας εξ αρχής τον αριθμό των ομάδων, είτε προσθέτοντας έναν αλγόριθμο προσδιορισμού του αριθμού των ομάδων. Αντίθετα, με την χρήση τεχνικών επεξεργασίας εικόνας είναι περισσότερο περίπλοκο να χωρίσεις τα εικονοστοιχεία σε έναν προκαθορισμένο αριθμό ομάδων αν η φύση του προβλήματος το απαιτεί. Τέτοιες απαιτήσεις έχει το πρόβλημα της κατάτμησης

των εικόνων των μικροσυστοιχιών, το οποίο από την φύση του σκοπεύει στον διαχωρισμό των εικονοστοιχείων σήματος από τα εικονοστοιχεία υποβάθρου.

Στην βιβλιογραφία όλες οι μέθοδοι που είναι βασισμένες σε τεχνικές ομαδοποίησης χρησιμοποιούν προκαθορισμένο αριθμό ομάδων ίσο με δύο [160, 162, 167], με σκοπό να διαχωρίσουν τα εικονοστοιχεία υποβάθρου με τα εικονοστοιχεία σήματος. Η παρούσα διδακτορική διατριβή εισήγαγε την τρίτη ομάδα με σκοπό σε αυτήν να καταχωρηθούν εικονοστοιχεία που πρέπει να εξαιρεθούν από την διαδικασία της εξαγωγής των ποσοτικοποιημένων τιμών. Παρακάτω περιγράφονται αναλυτικά οι τρεις κατηγορίες από εικονοστοιχεία, στις οποίες θα τα διαχωρίσει η κατάτμηση, α) τα εικονοστοιχεία σήματος, β) τα εικονοστοιχεία υποβάθρου και γ) τα εικονοστοιχεία τεχνουργημάτων. Κατά την υλοποίηση της μεθοδολογίας υλοποιήθηκαν και προβλήματα δύο κατηγοριών (εικονοστοιχεία σήματος και εικονοστοιχεία υποβάθρου) με σκοπό την ανάδειξη της χρησιμότητας της τρίτης κατηγορίας.

Εικονοστοιχεία σήματος: Τα εικονοστοιχεία σήματος είναι τα εικονοστοιχεία που αντιπροσωπεύουν την ποσοτικοποίηση της υβριδοποίησης κάθε ανιχνευτή με τα δείγματα. Η μέση φωτεινότητα δηλαδή αυτών των εικονοστοιχείων είναι αυτή που κυρίως θα συνεισφέρει στις ποσοτικοποιημένες τιμές που θα προκύψουν από την διαδικασία της εξαγωγής της έντασης για κάθε μια κηλίδα. Πρακτικά, τα εικονοστοιχεία σήματος είναι τα εικονοστοιχεία που βρίσκονται μέσα στα όρια της κηλίδας. Το βασικό χαρακτηριστικό των εικονοστοιχείων αυτών είναι η υψηλή φωτεινότητά τους. Αυτό βέβαια συμβαίνει όταν ανήκουν σε κάποια κηλίδα όπου ο ανιχνευτής που της αντιστοιχεί έχει υβριδοποιηθεί με τουλάχιστον ένα από τα δείγματα. Αντίθετα τα εικονοστοιχεία σήματος μιας μη-υβριδοποιημένης κηλίδας έχουν χαμηλή φωτεινότητα. Στις περισσότερες περιπτώσεις βέβαια, ακόμα και για κηλίδες με μικρό βαθμό υβριδοποίησης, τα εικονοστοιχεία σήματος έχουν υψηλότερη φωτεινότητα από το υπόβαθρο της εικόνας. Ένα άλλο χαρακτηριστικό που έχουν τα εικονοστοιχεία αυτά είναι η κοντινή απόσταση που έχουν από το κέντρο της κηλίδας ή του κελιού που απομονώνει μια κηλίδα.

Εικονοστοιχεία υποβάθρου: Τα εικονοστοιχεία υποβάθρου είναι τα εικονοστοιχεία, τα οποία χαρακτηρίζουν τον θόρυβο υποβάθρου της περιοχής γύρω από την κηλίδα. Ενώ θα αναμενόταν τα εικονοστοιχεία αυτά να έχουν μηδενική φωτεινότητα, αυτό δεν συμβαίνει. Φαινόμενα εκπομπής του ίδιου του υποστρώματος αλλά και φαινόμενα ανάκλασης κατά την διάρκεια της σάρωσης αυξάνουν σε μικρό βαθμό την φωτεινότητα τους, δημιουργώντας θόρυβο υποβάθρου στην εικόνα. Η μέση ή η διάμεσος τιμή των εικονοστοιχείων υποβάθρου αφαιρείται από την μέση τιμή των εικονοστοιχείων σήματος με σκοπό να προκύψουν οι διορθωμένες τιμές. Τα φαινόμενα της εκπομπής και της ανάκλασης από το υπόστρωμα μπορεί να συμβαίνουν τοπικά σε μια μικροσυστοιχία, γεγονός που σημαίνει ότι δεν αρκεί να εντοπιστούν τα εικονοστοιχεία υποβάθρου από ένα σημείο της εικόνας και αυτά να χρησιμοποιηθούν για την διόρθωση των τιμών όλων των κηλίδων. Πρέπει δηλαδή για κάθε μια κηλίδα να εντοπίζονται και τα εικονοστοιχεία υποβάθρου μέσα στο κελί της. Τα εικονοστοιχεία υποβάθρου βρίσκονται έξω από τα όρια της κηλίδας, γεγονός που σημαίνει ότι η απόσταση τους από το κέντρο του κελιού είναι εν γένει μεγαλύτερη από αυτήν των εικονοστοιχείων σήματος.

Εικονοστοιχεία τεχνουργημάτων: Την τρίτη κλάση εικονοστοιχείων την εισάγαμε με σκοπό να κατηγοριοποιήσουμε τα εικονοστοιχεία, οι φωτεινότητες των οποίων δεν θα πρέπει να συμπεριληφθούν στους υπολογισμούς των διορθωμένων τιμών και κατ' επέκταση στους υπολογισμούς των λόγων μεταξύ των δύο καναλιών. Την ονομάσαμε έτσι επειδή τα εικονοστοιχεία που συμπεριλαμβάνονται σε αυτήν είναι κυρίως εικονοστοιχεία από τεχνουργήματα που υπάρχουν στην εικόνα. Τα τεχνουργήματα αυτά μπορεί να προέρχονται είτε από μολύνσεις πάνω στην μικροσυστοιχία, είτε από γρατζουνιές στην επιφάνεια του υποστρώματος, είτε από σταγόνες των φθοριζόντων δειγμάτων σε σημεία εκτός των θέσεων των ανιχνευτών. Τα εικονοστοιχεία τεχνουργημάτων είναι εικονοστοιχεία υψηλής φωτεινότητας τα οποία πιθανότατα θα συμπεριλαμβάνονταν στα εικονοστοιχεία σήματος, αλλοιώνοντας έτσι τις ποσοτικοποιημένες τιμές. Πέραν των εικονοστοιχείων των τεχνουργημάτων η κατηγορία αυτή περιέχει και τα εικονοστοιχεία των σκοτεινών οπών που πολλές φορές βρίσκονται μέσα στις κηλίδες με σχήμα ντόνατς (donut spots) [191]. Οι οπές αυτές δημιουργούνται από ελλειπή και ανομοιογενή υβριδοποίηση μεταξύ του ανιχνευτή και του δείγματος. Τέλος η κατηγορία αυτή περιέχει εικονοστοιχεία από το περίγραμμα της κηλίδας. Τα τρία είδη εικονοστοιχείων που έχει

αυτή η κατηγορία συνήθως έχουν παρόμοιες φωτεινότητες και αυτός είναι ο λόγος που τα συμπεριλάβαμε σε μια ενιαία κατηγορία. Για παράδειγμα τα εικονοστοιχεία των σκοτεινών οπών και τα εικονοστοιχεία του περιγράμματος της κηλίδας είναι χαμηλότερης φωτεινότητας από τα εικονοστοιχεία σήματος, ωστόσο είναι επηρεασμένα από την φωτεινότητα του σήματος που είναι πολύ κοντά τους και συνεπώς η φωτεινότητα τους είναι κατά πολύ υψηλότερη των εικονοστοιχείων υποβάθρου. Με απλά λόγια, μπορούμε να αναφέρουμε ότι η τρίτη κατηγορία αποτελεί μια δεξαμενή εικονοστοιχείων, τα οποία δεν πρέπει να συμπεριληφθούν στην διαδικασία της εξαγωγής των ποσοτικοποιημένων τιμών, γεγονός που δεν συμβαίνει με την χρήση δύο ομάδων. Στην περίπτωση των δύο ομάδων τα εικονοστοιχεία αυτά συμπεριλαμβάνονται κατά περίπτωση είτε στα εικονοστοιχεία σήματος, είτε στα εικονοστοιχεία υποβάθρου, αλλοιώνοντας της ποσότητες που εξάγονται.

5.4.1. Κατάτμηση χωρίς επίβλεψη

Οι κατάτμηση χωρίς επίβλεψη είναι βασισμένη σε τεχνικές ομαδοποίησης. Τα εικονοστοιχεία με παρόμοια χαρακτηριστικά ομαδοποιούνται σε δύο οι τρεις ομάδες. Οι αλγόριθμοι που χρησιμοποιήθηκαν είναι οι περισσότερο διαδεδομένοι αλγόριθμοι για προβλήματα ομαδοποίησης και έχουν χρησιμοποιηθεί στο παρελθόν για την κατάτμηση εικόνων μικροσυστοιχιών, αλγόριθμο K-μέσων (K-means) [152] και τον ασαφή αλγόριθμο K-μέσων (Fuzzy C means - FCM) [161]. Ωστόσο, η προτεινόμενη μέθοδος αφενός χρησιμοποιεί το καινοτόμο σύνολο χαρακτηριστικών, και αφετέρου εισάγει και την τρίτη ομάδα εικονοστοιχείων με σκοπό να καταχωρηθούν εκεί όλα τα εικονοστοιχεία τεχνουργημάτων. Διάφορες παραλλαγές των δύο αλγορίθμων, του αριθμού των ομάδων, και του συνόλου των χαρακτηριστικών υλοποιήθηκαν.

5.4.1.1. Ομαδοποίηση με τον αλγόριθμο K-μέσων

Η πρώτη προσέγγιση που υλοποιήθηκε είναι βασισμένη στον αλγόριθμο ομαδοποίησης K-μέσων. Ο αλγόριθμος K-μέσων χρησιμοποιεί ένα κριτήριο τετραγωνικού σφάλματος, το οποίο και υπολογίζεται για κάθε μια ομάδα. Το κριτήριο τετραγωνικού σφάλματος για την ομαδοποίηση σε τρεις ομάδες δίνεται από της σχέση:

$$E_{Tot}^2 = \sum_{k=1}^3 e_k^2, \quad (5.6)$$

Όπου

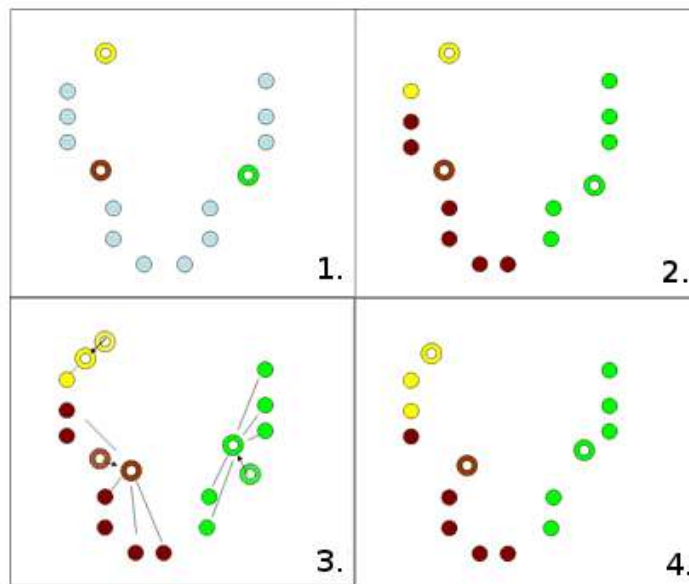
$$E_k^2 = \sum_{i=1}^{n_k} (\overline{f_{ik}} - M_k)^2, \quad k = 1, 2, 3 \quad (5.7)$$

και $\overline{f_{ik}}$ είναι το διάνυσμα χαρακτηριστικών του ι-οστού εικονοστοιχείου, n_k είναι ο αριθμός των εικονοστοιχείων που ανήκουν στην K-οστή ομάδα και M_k είναι τα κεντροειδή της K-οστής ομάδας αντιστοίχως:

$$M_k = \left(\frac{1}{n_k} \right) \sum_{i=1}^{n_k} \overline{f_{ik}}. \quad (5.8)$$

Σημειώστε πως οι αρχικές τιμές για τα κεντροειδή ανατίθενται τυχαία. Η ανάθεση θα μπορούσε να γίνει είτε τυχαία είτε χρησιμοποιώντας τις τιμές των K πρώτων στοιχείων. Σε κάθε επανάληψη του αλγορίθμου κάθε εικονοστοιχείο ανατίθεται σε μια από τις ομάδες και στην συνέχεια τα κεντροειδή της κάθε ομάδας επαναπροσδιορίζονται συμπεριλαμβάνοντας και τα νέα εικονοστοιχεία που έχουν ανατεθεί σε αυτήν την επανάληψη. Τα βήματα του αλγορίθμου κατά την διάρκεια μιας επανάληψης φαίνονται στο σχήμα 5.4.

Ο αλγόριθμος συνήθως τελειώνει όταν ικανοποιείται ένα κριτήριο σύγκλισης. Το κριτήριο σύγκλισης θα μπορούσε να βασιστεί στο τετραγωνικό σφάλμα, αλλά αυτό δεν είναι απαραίτητο. Για παράδειγμα, ο αλγόριθμος θα μπορούσε να τερματίζει όταν δεν υπάρχει κανένα εικονοστοιχείο (ή υπάρχουν ελάχιστα) που να ανατίθεται σε διαφορετικές ομάδες. Άλλες τεχνικές τερματισμού απλά τερματίζουν μετά από ένα συγκεκριμένο πλήθος επαναλήψεων. Μπορεί να προβλεφτεί ένας μέγιστος αριθμός επαναλήψεων που να εγγυάται τον τερματισμό του αλγορίθμου ακόμα και αν δεν συγκλίνει.



Σχήμα 5.4: Τα βήματα που εκτελεί επαναληπτικά ο αλγόριθμος K-μέσων

5.4.2.Ομαδοποίηση με τον Ασαφή αλγόριθμο K-μέσων

Μερικές από τις μεθόδους για την κατάτμηση των μικροσυστοιχιών, αντί να χρησιμοποιήσουν τον αλγόριθμο K-μέσων, ο οποίος είναι ένας σαφής αλγόριθμος που αναθέτει εξ' ολοκλήρου κάθε εικονοστοιχείο σε μία και μόνο ομάδα, χρησιμοποιούν τον ασαφή αλγόριθμο C-μέσων (Fuzzy C-means ή FCM). Ο ασαφής αλγόριθμος C-μέσων είναι μια τεχνική ομαδοποίησης κατά την οποία το σύνολο δεδομένων (στην προκειμένη περίπτωση τα εικονοστοιχεία μέσα στο κελί της κηλίδας) ομαδοποιούνται σε ομάδες, όπου το κάθε εικονοστοιχείο ανήκει σε κάθε μια ομάδα με ένα βαθμό συμμετοχής. Για να γίνει αυτό κατανοητό ας υποθέσουμε στο δικό μας πρόβλημα ότι ένα πολύ φωτεινό εικονοστοιχείο που είναι κοντά στο κέντρο του κελιού θα έχει μεγάλο βαθμό συμμετοχής στην ομάδα των εικονοστοιχείων σήματος, ενώ μικρότερους βαθμούς συμμετοχής στις ομάδες υποβάθρου και τεχνουργημάτων.

Σε αντιστοιχία με τον αλγόριθμο K-μέσων ο οποίος ελαχιστοποιεί το τετραγωνικό σφάλμα, έτσι και ο ασαφής αλγόριθμος C-μέσων είναι βασισμένος στην ελαχιστοποίηση μιας αντικειμενικής συνάρτησης όπως αυτή δίνεται από την σχέση:

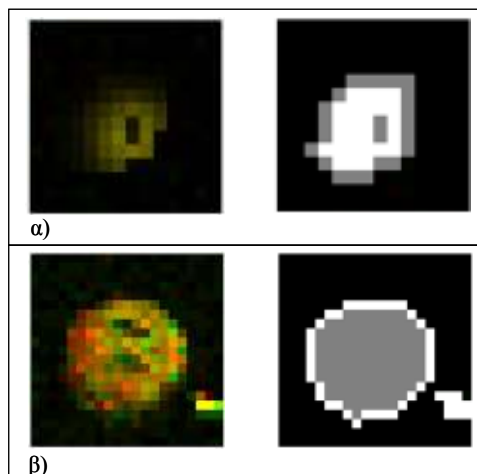
$$J_m = \sum_i \sum_{k=1}^{n_k} (u_{ik})^\gamma \left\| \overline{f_{ik}} - M_k \right\|^2, \quad (5.9)$$

όπου, x_{ik} είναι το διάνυσμα χαρακτηριστικών του i -οστού εικονοστοιχείου, M_k είναι το κεντροειδές της κάθε ομάδας, u_{ik} είναι ο βαθμός συμμετοχής του εικονοστοιχείου με διάνυσμα χαρακτηριστικών x_{ik} για κάθε ομάδα, $\left\| \overline{f_{ik}} - M_k \right\|^2$ είναι η Ευκλείδεια απόσταση μεταξύ του διανύσματος χαρακτηριστικών x_{ik} και του κεντροειδούς M_k , n_k είναι ο αριθμός των εικονοστοιχείων που ανήκουν στην K -οστή ομάδα. Η παράμετρος γ είναι το εκθετικό βάρος του u_{ik} , ο οποίος ελέγχει την ασάφεια της εξαγόμενης ομάδας. Σύμφωνα με τα παραπάνω το κάθε εικονοστοιχείο αντιστοιχίζεται στην κατηγορία για την οποία έχει τον μεγαλύτερο βαθμό συμμετοχής.

5.4.2.1. Αντιστοίχιση ομάδων με κατηγορία

Η ομαδοποίηση των εικονοστοιχείων δεν ολοκληρώνει την κατάτμηση των κηλίδων. Αυτό που επιτυγχάνει μέχρι στιγμής είναι η ομαδοποίηση των εικονοστοιχείων σε δύο ή τρεις ομάδες, χωρίς όμως να έχουν αντιστοιχηθεί αυτές οι ομάδες με το υπόβαθρο, το σήμα ή τα τεχνουργήματα στην εικόνα. Για να ολοκληρωθεί η κατάτμηση της εικόνας απαιτείται η αντιστοίχιση των παραγόμενων ομάδων εικονοστοιχείων με τις κατηγορίες εικονοστοιχείων που υπάρχουν στην εικόνα. Η διαδικασία αυτή όπως εφαρμόζεται σε ήδη υπάρχουσες μεθόδους [154-155, 157], γίνεται με την χρήση απλών κανόνων. Συγκεκριμένα εφαρμόζεται ο κανόνας της υψηλότερης μέσης φωτεινότητας της ομάδας. Η ομάδα δηλαδή η οποία έχει την υψηλότερη μέση φωτεινότητα χαρακτηρίζεται ως σήμα ενώ η ομάδα με την χαμηλότερη μέση φωτεινότητα ως υπόβαθρο. Ο κανόνας αυτός παρουσιάζει σχεδόν απόλυτη ακρίβεια όταν οι μέθοδοι πραγματοποιούν την κατάτμηση για δύο κατηγορίες εικονοστοιχείων. Είναι πολύ σπάνιο φαινόμενο τα εικονοστοιχεία του υποβάθρου να έχουν μεγαλύτερη μέση φωτεινότητα από αυτά του σήματος. Ωστόσο, η χρήση της τρίτης ομάδας περιπλέκει αρκετά αυτήν την διαδικασία. Στο σχήμα 5.5 φαίνονται τα αποτελέσματα της εφαρμογής του κανόνα υψηλότερης φωτεινότητας για δύο ενδεικτικές κηλίδες κατά την ομαδοποίηση των εικονοστοιχείων της εικόνας σε τρεις ομάδες. Είναι αρκετές οι φορές όπου στην τρίτη ομάδα εμπεριέχεται ένα τεχνουργήμα

υψηλής φωτεινότητας, γεγονός το οποίο προσδίδει στην τρίτη ομάδα μέση τιμή φωτεινότητας μεγαλύτερη από εκείνη που έχει η ομάδα του σήματος. Το αποτέλεσμα της κατάτμησης για μια τέτοια περίπτωση φαίνεται στο σχήμα 5.5β.



Σχήμα 5.5: Αποτέλεσμα της κατάτμησης που φανερώνει την αδυναμία του κανόνα υψηλότερης φωτεινότητας να αντιστοιχίσει σωστά τις παραγόμενες ομάδες με τις κατηγορίες εικονοστοιχείων: α) αποτέλεσμα σωστής αντιστοίχισης, β) αποτέλεσμα λανθασμένης αντιστοίχισης.

Όπως παρατηρείται στο σχήμα 5.5β το τεχνούργημα το οποίο βρίσκεται στην κάτω-δεξιά περιοχή του κελίου, έχει υψηλότερη φωτεινότητα από τα εικονοστοιχεία της ίδιας της κηλίδας. Αυτό έχει ως αποτέλεσμα τα εικονοστοιχεία του τεχνουργήματος και τα εικονοστοιχεία του περιγράμματος της κηλίδας, που αποτελούν την ομάδα των τεχνουργημάτων, να κατηγοριοποιούνται ως εικονοστοιχεία σήματος, ενώ τα εικονοστοιχεία της κηλίδας, τα οποία έχουν την δεύτερη κατά σειρά μεγαλύτερη μέση φωτεινότητα από τις τρεις ομάδες κατηγοριοποιούνται ως εικονοστοιχεία τεχνουργήματος. Η αδυναμία αυτή της διαδικασίας που είναι βασισμένη σε κανόνες σαφέστατα μπορεί να βελτιωθεί με την προσθήκη νέων κανόνων. Ωστόσο αναδεικνύει το γεγονός ότι η ομαδοποίηση κατηγοριοποιεί έμμεσα το κάθε εικονοστοιχείο σε μια κατηγορία, μέσω της κατηγοριοποίησης της ομάδας στην οποία ανήκει. Το γεγονός αυτό έδωσε το έναυσμα για την υλοποίηση μιας διαφορετικής προσέγγισης για την κατάτμηση των εικόνων των μικροσυστοιχιών, στην οποία το κάθε εικονοστοιχείο θα ταξινομείται απευθείας σε μία από τις κατηγορίες.

5.4.3. Κατάτμηση με επίβλεψη

Σε αυτό το εδάφιο περιγράφεται η προτεινόμενη μεθοδολογία για την κατάτμηση των εικόνων των μικροσυστοιχιών με επίβλεψη. Η πρωτοτυπία της μεθοδολογίας αυτής έγκειται στο γεγονός ότι είναι βασισμένη σε τεχνικές ταξινόμησης. Η βασική ιδέα είναι η ταξινόμηση όλων των εικονοστοιχείων της εικόνας σε δύο οι τρεις κατηγορίες, οι οποίες αντιπροσωπεύουν τα εικονοστοιχεία σήματος, τα εικονοστοιχεία υποβάθρου και τα εικονοστοιχεία τεχνουργημάτων στην εικόνα. Διάφορες παραλλαγές δοκιμάστηκαν είτε σε επίπεδο των χαρακτηριστικών που θα τροφοδοτήσουν τους ταξινομητές, είτε σε επίπεδο ταξινομητών. Πιο συγκεκριμένα έχουν χρησιμοποιηθεί διαφορετικά σύνολα χαρακτηριστικών αλλά και γραμμικών συνδυασμών αυτών με την χρήση της ανάλυσης κύριων συνιστωσών, ενώ κυρίως δύο διαφορετικοί ταξινομητές, ο Μπεϋζιανός ταξινομητής [192] αλλά και οι μηχανές διανυσμάτων υποστήριξης [193]. Η προτεινόμενη προσέγγιση αποδεικνύεται να είναι πολύ πιο αξιόπιστη από τις προσεγγίσεις, οι οποίες είναι βασισμένες σε τεχνικές ομαδοποίησης και οι οποίες έχουν επικρατήσει τα τελευταία χρόνια στην βιβλιογραφία. Αυτό συμβαίνει καθώς με την χρήση ταξινομητή κάθε εικονοστοιχείο της εικόνας άμεσα χαρακτηρίζεται σε ποια από τις κατηγορίες ανήκει. Αντίθετα όπως έχει ειπωθεί και ανωτέρω οι μέθοδοι κατάτμησης, οι οποίες είναι βασισμένες στις τεχνικές ομαδοποίησης, πρώτα ομαδοποιούν τα εικονοστοιχεία και την συνέχεια τα χαρακτηρίζουν έμμεσα χαρακτηρίζοντας την ομάδα την οποία ανήκουν, με χρήση απλών κανόνων. Οι κανόνες πολλές φορές αποτυγχάνουν να χαρακτηρίσουν σωστά τις ομάδες πλήττοντας έτσι και την αξιοπιστία των μεθόδων.

Ο μεγάλος αριθμός των εικονοστοιχείων τα οποία πρέπει να ταξινομηθούν απαιτούν έναν γρήγορο και αξιόπιστο ταξινομητή. Για τον λόγο αυτό ο πρώτος ταξινομητής που δοκιμάστηκε για να αντιμετωπίσει το συγκεκριμένο πρόβλημα ήταν ο Μπεϋζιανός ταξινομητής. Στην συνέχεια θέλοντας να αυξήσουμε την ακρίβεια της μεθόδου δοκιμάστηκαν πιο εξελιγμένοι ταξινομητές όπως οι γραμμικές και μη γραμμικές Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines - SVM).

5.4.3.1. Μπεϋζιανός Ταξινομητής

Η ιδέα του Μπεϋζιανού ταξινομητή είναι ο προσδιορισμός της a posteriori πιθανότητας το εικονοστοιχείο να ανήκει σε μια κλάση. Αυτό γίνεται με την βοήθεια του κανόνα του Μπέϋζ, ο οποίος δίνεται από την παρακάτω σχέση:

$$P(w_i | f) = \frac{p(f | w_i)P(w_i)}{\sum_{i=1}^2 p(f | w_i)P(w_i)} = \frac{p(f | w_i)P(w_i)}{p(f)}, \quad (5.10)$$

όπου, $f \in R^D$ είναι το διάνυσμα χαρακτηριστικών, D είναι η διάσταση του διανύσματος χαρακτηριστικών, $w_i : i = 1, 2$ είναι οι δύο κλάσεις, $P(w_i)$ είναι η a priori πιθανότητα ένα εικονοστοιχείο να ανήκει στην κλάση w_i , $P(w_i | f)$ είναι η υπό συνθήκη a posteriori πιθανότητα ένα εικονοστοιχείο να ανήκει σε μια κλάση, $p(f)$ είναι η πυκνότητα πιθανότητας όλων των εικονοστοιχείων, και $p(f | w_i)$ είναι η υπό συνθήκη πυκνότητα πιθανότητας όλων των εικονοστοιχείων που ανήκουν στην κλάση w_i .

Το θεώρημα μπορεί να εφαρμοστεί με διάφορες συναρτήσεις πυκνότητας πιθανότητας, ωστόσο αυτό εξαρτάται από την φύση των δεδομένων. Εμείς χρησιμοποιήσαμε μια Γκαουσιανή συνάρτηση πυκνότητας πιθανότητας η οποία δίνεται από την σχέση:

$$p(f) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} e^{-\frac{1}{2}(f-\mu_i)^T \Sigma_i^{-1} (f-\mu_i)}, \quad (5.11)$$

όπου μ_i και Σ_i είναι το διάνυσμα μέσων τιμών και ο πίνακας συνμεταβλητότητας των χαρακτηριστικών για την κλάση w_i αντίστοιχα, όπου δίνονται από τους τύπους:

$$\mu_i = \frac{1}{N_i} \sum_{f \in w_i} f, \quad \mu_i \in \mathfrak{R}^{11}, \quad (5.12)$$

Και

$$\Sigma_i = \frac{1}{N_i} \sum_{f \in w_i} f f^T - \mu_i \mu_i^T, \quad \Sigma_i \in \mathfrak{R}^{1 \times 11}, \quad (5.13)$$

Όπου N_i είναι των εικονοστοιχείων που ανήκουν στην κλάση w_i .

Κατά την φάση εκπαίδευσης προσδιορίζονται το διάνυσμα μέσω των τιμών και ο πίνακας συνμεταβλητότητας για κάθε μια από της κλάσης. Κατά την φάση ελέγχου, δοθέντων του διανύσματος μέσω των τιμών, του πίνακα συνμεταβλητότητας και της Γκαουσιανής συνάρτησης πυκνότητας πιθανότητας προσδιορίζεται η εκ των υστέρων πιθανότητα κάθε εικονοστοιχείου να ανήκει σε καθεμία από τις κλάσεις, σύμφωνα με την εξίσωση 5.10.

Ο Μπεϋζιανός ταξινομητής είναι εύκολα υλοποιήσιμος και σε αντίθεση με άλλους ταξινομητές απαιτείται μόνο ένα πέρασμα των δεδομένων εκπαίδευσης. Επίσης η ταξινόμηση κατά Μπέϋζ μπορεί εύκολα να χειριστεί ελλιπή δεδομένα απλά παραλείποντας την αντίστοιχη πιθανότητα όταν υπολογίζει τις πιθανοφάνειες του εικονοστοιχείου για κάθε κατηγορία. Από την άλλη πλευρά παρόλο που η προσέγγιση του Μπέϋζ είναι αρκετά απλή στην χρήση της δεν δίνει πάντα ικανοποιητικά αποτελέσματα. Αυτό συμβαίνει γενικά όταν τα χαρακτηριστικά δεν είναι ανεξάρτητα. Αυτός είναι και ο λόγος για τον οποίο εφαρμόστηκε η Ανάλυση Κύριων Συνιστωσών.

5.4.3.2. Μηχανές Διανυσμάτων Υποστήριξης

Οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines – SVM) είναι γραμμικές συναρτήσεις, στόχος των οποίων είναι η προσέγγιση ενός υπερεπιπέδου που διαχωρίζει ακριβέστερα τα δείγματα σε δύο κατηγορίες (θεωρώντας ότι το πρόβλημα περιλαμβάνει δύο κατηγορίες), μεγιστοποιώντας το περιθώριο στο χώρο των χαρακτηριστικών. Οι Μηχανές Διανυσμάτων Υποστήριξης αναζητούν ένα υποσύνολο των εικονοστοιχείων, το οποίο δημιουργεί τα όρια μεταξύ των δύο κατηγοριών (διανύσματα υποστήριξης). Ας υποθέσουμε το σύνολο εκπαίδευσης για δύο κατηγορίες $\{(\vec{f}_1, c_1^*), (\vec{f}_2, c_2^*), \dots, (\vec{f}_i, c_i^*)\}$, όπου τα $\{\vec{f}_1, \vec{f}_2, \dots, \vec{f}_i\} \in \mathbb{R}^2$ είναι τα διανύσματα

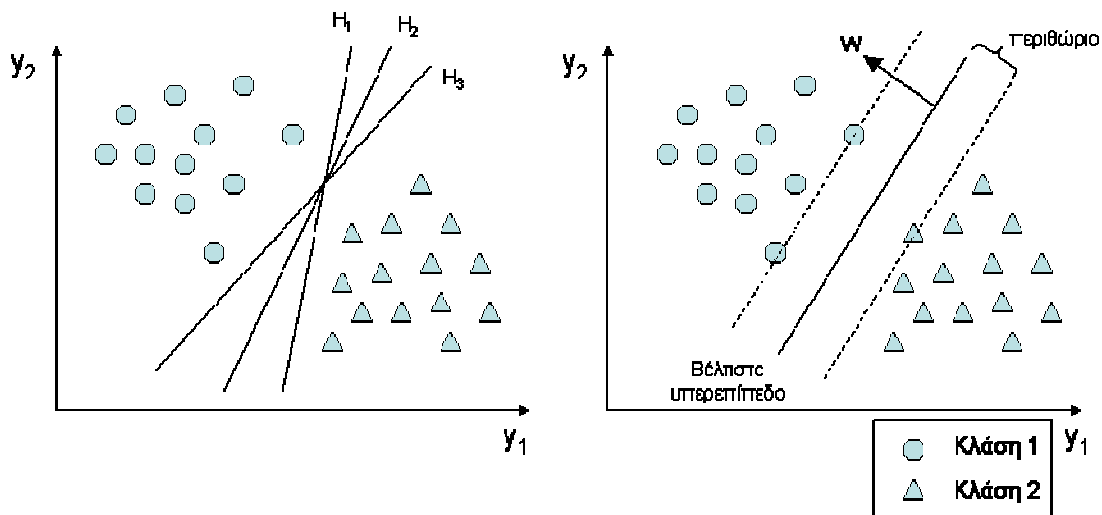
χαρακτηριστικών για κάθε εικονοστοιχείο, c_i^* είναι η κατηγορία στην οποία ανήκει το εικονοστοιχείο με διάνυσμα χαρακτηριστικών f_i . και είναι είτε 0 (για το υπόβαθρο) είτε 1 (για το σήμα). Το υπερεπίπεδο το οποίο χωρίζει τα διανύσματα χαρακτηριστικών στον χώρο δίνεται από την σχέση:

$$\bar{w}x - b = 0, \tag{5.14}$$

όπου το διάνυσμα \bar{w} είναι το κάθετο διάνυσμα στο υπερεπίπεδο, ενώ το b ορίζει την μετατόπιση του υπερεπιπέδου από την αρχή του διανύσματος \bar{w} .

Το σχήμα 5.6 παρουσιάζει ένα δισδιάστατο παράδειγμα κατά το οποίο οι μηχανές διανυσμάτων υποστήριξης επιλέγουν το υπερεπίπεδο (στον δισδιάστατο χώρο είναι μια γραμμή), το οποίο μεγιστοποιεί το περιθώριο μεταξύ των δύο κατηγοριών.

Στο παράδειγμα φαίνεται ότι πολλές γραμμές (H_1, H_2, H_3) μπορούν να χωρίσουν το σύνολο εκπαίδευσης, οι μηχανές διανυσμάτων υποστήριξης όμως επιλέγουν την γραμμή, η οποία μεγιστοποιεί το περιθώριο όπως φαίνεται στο διάγραμμα στα δεξιά του σχήματος 5.6.



Σχήμα 5.6: Εύρεση του βέλτιστου υπερεπιπέδου για των διαχωρισμό των προτύπων από τις μηχανές διανυσμάτων υποστήριξης. Αριστερά φαίνονται τρία υπερεπίπεδα που χωρίζουν μπορούν να χωρίσουν τα πρότυπα, ενώ δεξιά φαίνεται το υπερεπίπεδο το οποίο μεγιστοποιεί το περιθώριο.

Για προβλήματα που δεν είναι γραμμικά διαχωρίσιμα, μπορούν να χρησιμοποιηθούν μέθοδοι πυρήνων που μετασχηματίζουν ένα μη γραμμικό χώρο εισόδου σε ένα γραμμικό χώρο χαρακτηριστικών. Είναι δηλαδή ένας μη γραμμικός μετασχηματισμός από το χώρο εκπαίδευσης των χαρακτηριστικών των εικονοστοιχείων, σε ένα χώρο χαρακτηριστικών μεγαλύτερης διάστασης κατασκευάζοντας το διαχωριστικό υπερεπίπεδο στον νέο χώρο. Κάποιοι γνωστοί πυρήνες είναι ο πολυωνυμικός πυρήνας, ο πυρήνας ακτινικών συναρτήσεων βάσης και ο σιγμοειδής πυρήνας. Για τον διαχωρισμό των εικονοστοιχείων χρησιμοποιήθηκε ο πυρήνας ακτινικών συναρτήσεων βάσης.

Η δυική μορφή των μηχανών διανυσμάτων υποστήριξης αναδεικνύει ότι ο διαχωρισμός των προτύπων από το υπερεπίπεδο είναι συνάρτηση μόνο των διανυσμάτων υποστήριξης. Η δυική μορφή δίνεται από την σχέση:

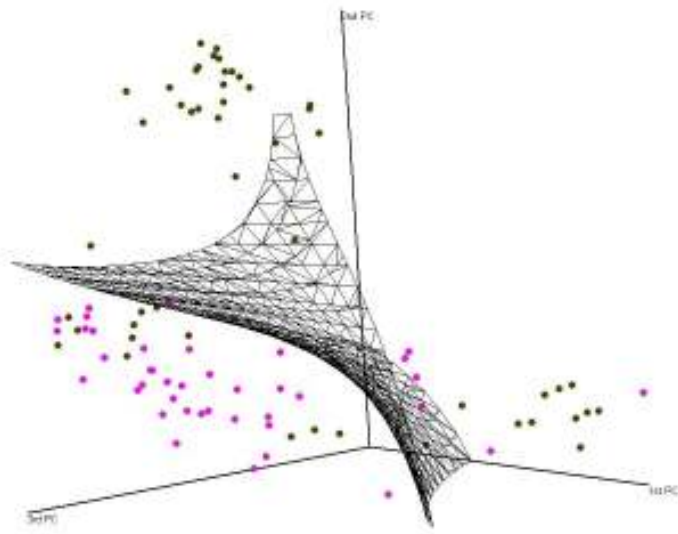
$$\max_a W(a) = \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m c_i c_j k(\vec{f}_i, \vec{f}_j) a_i a_j, \quad a_i \in [0, C], \quad (5.15)$$

$$\sum_{i=1}^m c_i a_i = 0, \quad (5.16)$$

Όπου m είναι ο αριθμός των διανυσμάτων των χαρακτηριστικών \vec{f}_i , C είναι παράμετρος των μηχανών διανυσμάτων υποστήριξης, ενώ οι όροι a_i είναι η δυική αναπαράσταση του διανύσματος των βαρών και οι οποίοι βελτιστοποιούνται κατά την διάρκεια της εκπαίδευσης. Τέλος $kn(\vec{f}_i, \vec{f}_j)$ είναι η συνάρτηση του πυρήνα όπου στην περίπτωση είναι ο πυρήνας ακτινικών συναρτήσεων βάσης και δίνεται από την σχέση:

$$kn(\vec{f}_i, \vec{f}_j) = e^{-\gamma |\vec{f}_i - \vec{f}_j|^2}. \quad (5.17)$$

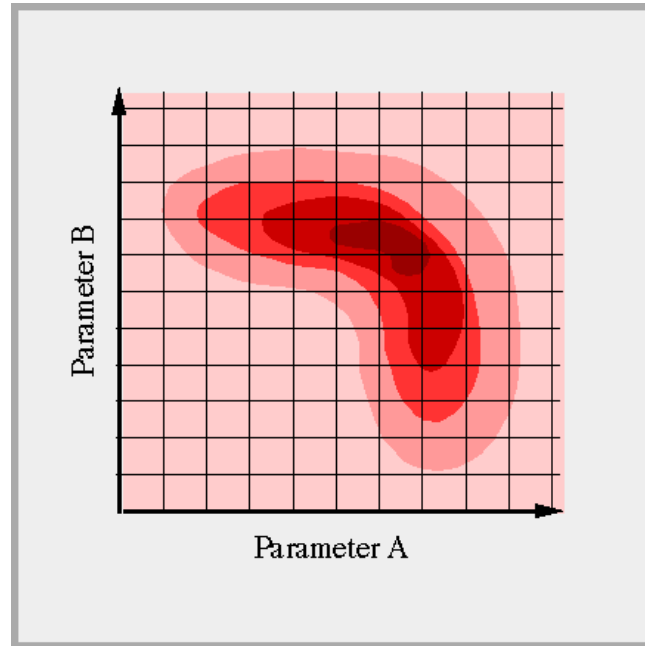
Όπου γ είναι και αυτή παράμετρος του πυρήνα. Το σχήμα 5.7 αναπαριστά ένα παράδειγμα μη γραμμικών διανυσμάτων υποστήριξης.



Σχήμα 5.7: Παράδειγμα μη γραμμικών μηχανών διανυσμάτων υποστήριξης στον τρισδιάστατο χώρο.

Η επιτυχία της ταξινόμησης με τις μηχανές διανυσμάτων υποστήριξης εξαρτάται σε μεγάλο βαθμό από την βελτιστοποίηση των παραμέτρων C και γ . Για τον λόγο αυτό εφαρμόσαμε την τεχνική της δικτυωτής αναζήτησης (Grid Search), για την βελτιστοποίηση των παραμέτρων. Πρακτικά η μέθοδος αυτή εξετάζει την ακρίβεια τις μεθόδου για όλα τα πιθανά ζεύγη των δύο παραμέτρων επιλέγοντας το ζεύγος για το οποίο η μέθοδος γίνεται πιο αξιόπιστη. Στο σχήμα 5.8 απεικονίζεται ένα παράδειγμα της τεχνικής της δικτυωτής αναζήτησης. Από την εφαρμογή της, εξάχθηκε ότι για το πρόβλημα δύο κλάσεων (κατηγοριοποίηση των εικονοστοιχείων σε εικονοστοιχεία σήματος και υποβάθρου) το βέλτιστο ζεύγος παραμέτρων C και γ είναι το (16,1), ενώ για το πρόβλημα τριών κλάσεων (εικονοστοιχεία σήματος, υποβάθρου και τεχνουργημάτων), το βέλτιστο ζεύγος είναι (16,2).

Ο σχεδιασμός των Μηχανών Διανυσμάτων Υποστήριξης αφορά προβλήματα που έχουν δύο μόνο κατηγορίες, αλλά χρησιμοποιώντας τεχνικές όπως ένας-εναντίον-ενός και ένας-εναντίον-όλων είναι δυνατή η εφαρμογή τους σε προβλήματα με περισσότερες κατηγορίες. Ένα άλλο μειονέκτημα είναι η ευαισθησία σε μη ισορροπημένες κατηγορίες, δηλαδή σε περιπτώσεις που ο αριθμός προτύπων κάθε κατηγορίας διαφέρει σημαντικά



Σχήμα 5.8: Η τεχνική δικτυωτής αναζήτησης για την βελτιστοποίησης δύο παραμέτρων. Η διαβαθμίσεις στο χρώμα του διαγράμματος αντιστοιχούν στην ακρίβεια της μεθόδου. Όσο πιο σκούρο είναι το χρώμα τόσο η μέθοδος παρουσιάζει μεγαλύτερη αξιοπιστία. Το ζεύγος που θα επιλέξει η μέθοδος για τις παραμέτρους θα είναι το ζεύγος το οποίο βρίσκεται στο πιο σκούρο σημείο του διαγράμματος.

5.5.Σύνολα δεδομένων

Η συγκεκριμένη μεθοδολογία αξιολογήθηκε χρησιμοποιώντας δύο σύνολα δεδομένων. Πιο συγκεκριμένα χρησιμοποιήθηκαν και πραγματικές εικόνες από την βάση δεδομένων μικροσυστοιχιών του Στάνφορντ [74] αλλά και προσομοιωμένες εικόνες, οι οποίες δημιουργήθηκαν από τον προσομοίωση που πρότεινε ο Nykter και οι συνεργάτες [78]. Το πρόβλημα το οποίο αντιμετωπίστηκε όσον αφορά στην αξιολόγηση αυτής της μεθοδολογίας έγκειται στο γεγονός ότι οι μέθοδοι, οι οποίες είναι βασισμένες σε τεχνικές με επίβλεψη απαιτούν την επισημείωση των δειγμάτων. Στην προκειμένη περίπτωση δηλαδή απαιτείται η επισημείωση για κάθε ένα από τα εικονοστοιχεία της εικόνας. Όπως έχει ήδη αναφερθεί και ανωτέρω οι εικόνες της βάσης δεδομένων του Στάνφορντ είναι επισημειωμένες με την χρήση του προγράμματος ScanAlyze, η επισημείωση όμως αυτή δεν περιέχει πληροφορία για κάθε εικονοστοιχείο της εικόνας αλλά για κάθε μια κηλίδα. Για να εξάγουμε την απαιτούμενη πληροφορία για τις συγκεκριμένες εικόνες προσεγγίσαμε τον τρόπο με τον οποίο λειτούργησε το συγκεκριμένο πρόγραμμα με σκοπό

να εξάγει την επισημείωση. Όπως αναφέρθηκε και στο κεφάλαιο της ανασκόπησης της βιβλιογραφίας το συγκεκριμένο πρόγραμμα ήταν από τα πρώτα που κυκλοφόρησαν για την ανάλυση των εικόνων και έκανε χρήση της μεθόδου σταθερού κύκλου. Γνωρίζοντας λοιπόν από το αρχείο της επισημείωσης που μας παρέχει η βάση δεδομένων του Στάνφορντ, το κέντρο που έχει εντοπίσει την κάθε κηλίδα αλλά και την ακτίνα του κύκλου, εξήγαμε τον χαρακτηρισμό του κάθε εικονοστοιχείου το οποίο θα μας έδινε το συγκεκριμένο πρόγραμμα. Όπως γίνεται αντιληπτό η βάση δεδομένων δεν μας δίνει κανένα στοιχείο για τυχόντα τεχνουργήματα στην εικόνα, και επομένως δεν μπορούμε στην περίπτωση των πραγματικών εικόνων να αξιολογήσουμε την μεθοδολογία μας για τρεις κατηγορίες εικονοστοιχείων. Αυτός είναι και ο βασικός λόγος, ο οποίος δίνει μεγάλη αξία στην χρησιμοποίηση προσομοιωμένων εικόνων για την αξιολόγηση της συγκεκριμένης μεθοδολογίας.

Δεδομένου ότι ο προσομοιωτής χρειάζεται μια σειρά δεδομένων ως είσοδο για να παράξει τις εικόνες, και με σκοπό οι εικόνες αυτές να προσεγγίζουν όσο το δυνατόν περισσότερο πραγματικές εικόνες μικροσυστοιχιών, χρησιμοποιήθηκαν τα δεδομένα του επισημειωμένου αρχείου από την βάση δεδομένων του Στάνφορντ ως είσοδο στον προσομοιωτή. Συγκεκριμένα χρησιμοποιήθηκαν οι συντεταγμένες της κάθε κηλίδας στην εικόνα αλλά και η μέση φωτεινότητα τους. Καταυτόν τον τρόπο αλλά και παραλλάσσοντας τις διάφορες παραμέτρους του προσομοιωτή δημιουργήσαμε εικόνες καλής, κανονικής, και κακής ποιότητας. Η κλιμάκωση της ποιότητας των προσομοιωμένων εικόνων συζητήθηκε στο 3^ο κεφάλαιο.

Ο προσομοιωτής δεν έχει προβλεφθεί να συμπεριλαμβάνει στην έξοδο του πληροφορίες για κάθε ένα από τα εικονοστοιχεία της εικόνας. Αυτό επιτευχθεί με συγκεκριμένες, στοχευμένες παρεμβάσεις/βελτιώσεις, οι οποίες έγιναν στον κώδικά του προσομοιωτή, ο οποίος ήταν ανοιχτός στο ευρύ κοινό. Σκοπός των βελτιώσεων αυτών ήταν να εξάγεται ταυτόχρονα με την δημιουργία των εικόνων και μία εικόνα, ίδιου μεγέθους με την παραγόμενη εικόνα, η οποία θα επισημαίνει σε ποια κατηγορία ανήκει το κάθε εικονοστοιχείο της εικόνας. Η εικόνα αυτή ξεκινάει με όλες τις τιμές μηδενικές και προσθέτει τιμές κατά την διάρκεια της προσομοίωσης. Προσθέτει 1 εκεί όπου ο

προσομοιωτής τοποθετεί κηλίδα και 2 όπου ο προσομοιωτής τοποθετεί κάποιο τεχνούργημα, ενώ όσα εικονοστοιχεία μένουν στην εικόνα αυτήν με μηδενική τιμή θα είναι και τα εικονοστοιχεία που θα αντιπροσωπεύουν το υπόβαθρο της εικόνας. Πέραν των τεχνουργημάτων στην ίδια κατηγορία τοποθετούνται και τα εικονοστοιχεία στο περίγραμμα της κάθε κηλίδας. Τα εικονοστοιχεία αυτά δεν πρέπει να συμπεριλαμβάνονται ως εικονοστοιχεία σήματος γιατί μειώνουν την μέση τιμή φωτεινότητας της κηλίδας, αλλά ούτε και ως εικονοστοιχεία υποβάθρου γιατί αυξάνουν την διάμεσο τιμή του υποβάθρου. Σε κάθε μια από της δυο αυτές περιπτώσεις η διορθωμένες τιμές, οι οποίες θα εξαχθούν και θα αντιπροσωπεύουν την ποσοτικοποίηση της υβριδοποίησης, θα είναι επηρεασμένες από τα εικονοστοιχεία αυτά.

5.6.Αποτελέσματα

Τα αποτελέσματα της κατάτμησης θα παρουσιαστούν παρακάτω για κάθε προσέγγιση ξεχωριστά. Αριθμητικά και σχηματικά αποτελέσματα έχουν εξαχθεί ξεχωριστά για την κατάτμηση με επίβλεψη και για την κατάτμηση χωρίς επίβλεψη. Τα αποτελέσματα αυτά παρουσιάστηκαν σε διάφορες εργασίες στα πλαίσια της διατριβής [180,186-189]. Στο τελευταίο εδάφιο του κεφαλαίου θα δούμε μια αναλυτική συγκριτική μελέτη μεταξύ των δυο προσεγγίσεων. Σκοπός των αποτελεσμάτων είναι η διαλεύκανση της συνεισφοράς των καινοτόμων σημείων που εισήγαγε η παρούσα διατριβή, όπως για παράδειγμα την εισαγωγή της τρίτης κατηγορίας εικονοστοιχείων, την εισαγωγή του καινοτόμου συνόλου των χαρακτηριστικών, αλλά και τέλος την χρήση τεχνικών ταξινόμησης για των διαχωρισμό των εικονοστοιχείων.

5.6.1.Κατάτμηση χωρίς επίβλεψη

Στην πρώτη φάση των αποτελεσμάτων μελετήθηκε η συμβολή της τρίτης ομάδας στο πρόβλημα, αλλά και η συνεισφορά των παραπάνω χαρακτηριστικών. Για τον λόγο αυτό υλοποιήθηκε η απλή προσέγγιση του αλγορίθμου K-μέσων με δύο ομάδες. Στον αντίποδα χρησιμοποιήθηκε ο ασαφής αλγόριθμος C-μέσων θέτοντας τις ομάδες ίσες με τρία. ($C=3$). Συγκεκριμένα υλοποιήθηκαν δύο μέθοδοι κατάτμησης βασισμένες στον αλγόριθμο K-μέσων για να κατατάξουν τα εικονοστοιχεία σε δύο κατηγορίες (εικονοστοιχεία σήματος και εικονοστοιχεία υποβάθρου). Η πρώτη [154-155] από αυτές χρησιμοποιεί μόνο τις

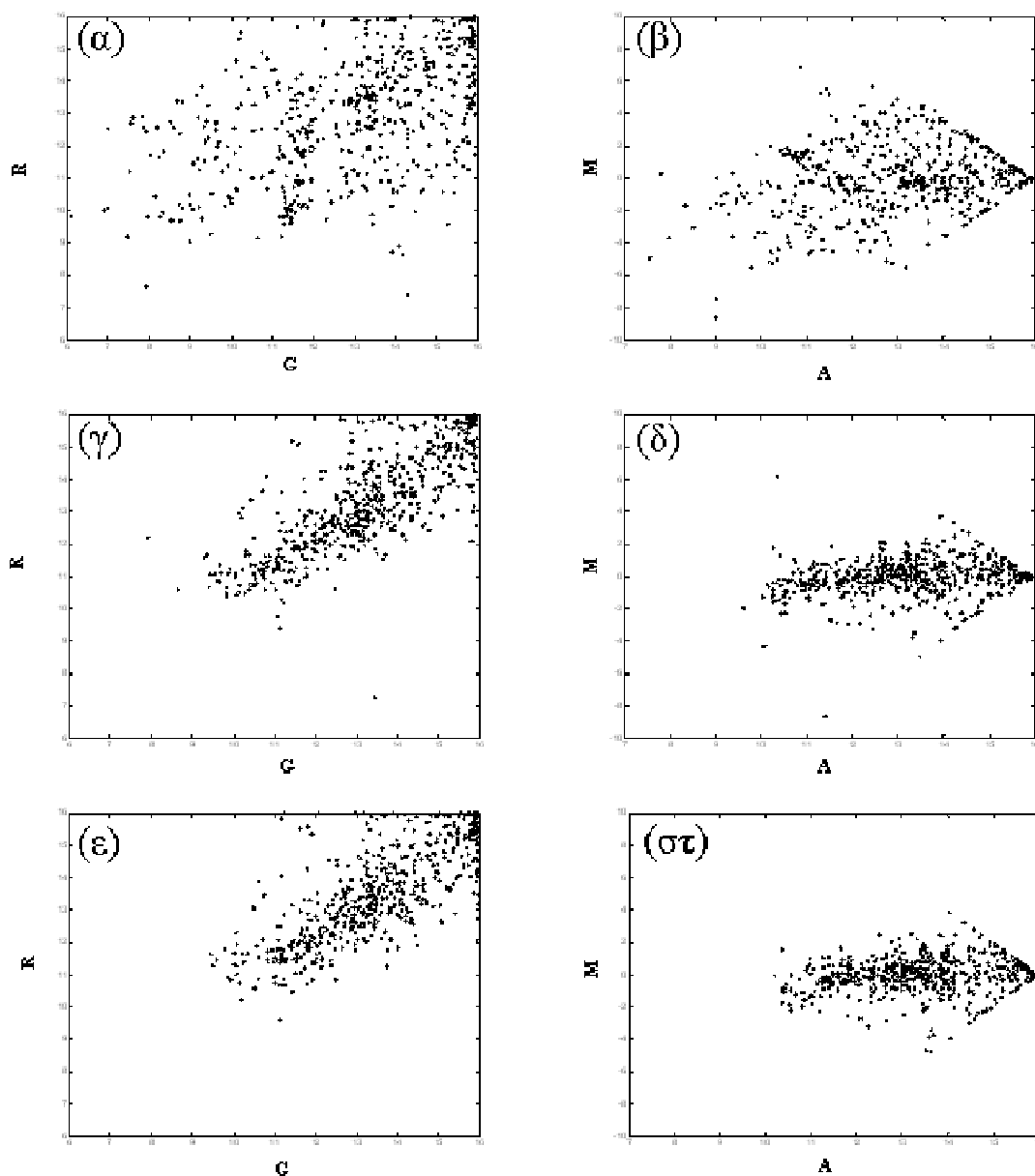
φωτεινότητες των δύο καναλιών ως χαρακτηριστικά των εικονοστοιχείων, ενώ η δεύτερη [156] χρησιμοποιεί εκτός από τις φωτεινότητες, τις συντεταγμένες του κάθε εικονοστοιχείου αλλά και την ευκλείδεια απόσταση τους από το κέντρο της κηλίδας. Η προτεινόμενη μέθοδος πρόσθεσε επίσης στα χαρακτηριστικά τις μέσες τιμές και τις τυπικές αποκλίσεις της 3X3 γειτονιάς του εικονοστοιχείου, καθώς επίσης και τον συντελεστή συσχέτισης της 11X11 γειτονιάς του εικονοστοιχείου με το Γκαουσιανό πρότυπο. Στο σχήμα 5.9 φαίνονται τα αποτελέσματα των τριών μεθόδων σε μία τυπική κηλίδα, σε μία κηλίδα με τεχνούργημα στο κελί της, και σε μία κηλίδα με σκοτεινή οπή στο κέντρο της (donut spot).

| | Εικόνα | Κ-μέσων 2 χαρακτ. 2 κλάσεις | Κ-μέσων 5 χαρακτ. 2 κλάσεις | Ασαφής C-μέσων 11 χαρακτ. 3 κλάσεις |
|-----------------------|--------|-----------------------------------|-----------------------------------|---|
| Τυπική κηλίδα | | | | |
| Κηλίδα «ντόνατ» | | | | |
| Κηλίδα με Τεχνούργημα | | | | |

Σχήμα 5.9: Τα αποτελέσματα της κατάτμησης των τριών υλοποιημένων μεθόδων για τρεις ενδεικτικές κηλίδες.

Τα αποτελέσματα των μεθόδων που υλοποιήθηκαν για δύο ομάδες εικονοστοιχείων ($K=2$) απεικονίζονται ως δυαδική εικόνα, όπου τα εικονοστοιχεία υποβάθρου είναι τα μαύρα εικονοστοιχεία ενώ τα εικονοστοιχεία σήματος τα άσπρα. Στο αποτέλεσμα της μεθόδου με τρεις ομάδες φαίνονται και περιοχές με γκρι εικονοστοιχεία, οι οποίες αντιπροσωπεύουν τα εικονοστοιχεία τεχνουργμάτων (τα εικονοστοιχεία δηλαδή της τρίτης κατηγορίας). Πρέπει να σημειώσουμε ότι για την αξιολόγηση της κατάτμησης χωρίς επίβλεψη χρησιμοποιήθηκαν εικόνες από την βάση του Στάνφορντ.

Σχολιάζοντας το παραπάνω σχήμα καταδεικνύεται η σημαντικότητα της χρήσης της τρίτης ομάδας. Ας παρατηρήσουμε τα αποτελέσματα για την κάθε μια κηλίδα ξεχωριστά. Η τυπική κηλίδα παρουσιάζεται να έχει καταταμθεί με αρκετά μεγάλη ακρίβεια και από τις τρεις μεθόδους. Η μόνη διαφοροποίηση φαίνεται στα αποτελέσματα της μεθόδου με τρεις ομάδες, η οποία έχει κατατάξει τα εικονοστοιχεία του περιγράμματος της κηλίδας στην τρίτη κατηγορία. Αυτό το γεγονός έχει νόημα διότι υπάρχουν εικονοστοιχεία του περιγράμματος που είναι χαμηλότερης φωτεινότητας από αυτά του σήματος, και τα οποία οι άλλες δύο μέθοδοι έχουν συμπεριλάβει είτε στα εικονοστοιχεία σήματος είτε στα εικονοστοιχεία υποβάθρου. Η μεγάλη διαφοροποίηση όμως που παρέχει η χρήση τρίτης ομάδας στις μεθόδους που είναι βασισμένες σε τεχνικές ομαδοποίησης φαίνεται στις άλλες δύο κηλίδες που χρησιμοποιήθηκαν. Η δεύτερη κηλίδα είναι μια κηλίδα με σκοτεινή οπή στο κέντρο της, η οποία έχει προκύψει από ελλειπή και ανομοιογενή υβριδοποίηση των δειγμάτων με τον ανιχνευτή. Παρατηρούμε ότι η προσέγγιση που χρησιμοποιεί ως χαρακτηριστικά μόνο τις φωτεινότητες των εικονοστοιχείων έχει εξαιρέσει τα εικονοστοιχεία της σκοτεινής οπής από τα εικονοστοιχεία σήματος. Αυτό το γεγονός πράγματι θα παράγει μια αρκετά ρεαλιστική τιμή για την μέση τιμή του σήματος της κηλίδας. Ωστόσο τα εικονοστοιχεία της σκοτεινής οπής έχουν καταταμθεί ως εικονοστοιχεία υποβάθρου, γεγονός που θα προκαλέσει αύξηση της διαμέσου τιμής του υποβάθρου κατά την εξαγωγή των ποσοτικοποιημένων τιμών. Αντίθετα, η προσέγγιση που χρησιμοποιεί και γεωμετρικά χαρακτηριστικά, κατατάσσει τα εικονοστοιχεία της σκοτεινής οπής στα εικονοστοιχεία σήματος. Σε αυτήν την περίπτωση η μέση τιμή των εικονοστοιχείων σήματος θα είναι ελαφρώς αυξημένη σε σχέση με την πραγματική. Είναι ξεκάθαρο από το σχήμα ότι η χρήση της τρίτης ομάδας έχει εξαιρέσει τα εικονοστοιχεία της σκοτεινής οπής και από τις δύο ομάδες από τις οποίες θα υπολογιστούν οι διορθωμένες τιμές της έντασης. Τέλος, παρόμοιο φαινόμενο παρατηρείται για τα εικονοστοιχεία των δύο τεχνουργημάτων στην τρίτη ενδεικτική κηλίδα που χρησιμοποιήσαμε. Τα εικονοστοιχεία των τεχνουργημάτων κατατάσσονται με μεγάλη ακρίβεια στην τρίτη ομάδα, σε αντίθεση με τις προσεγγίσεις που χρησιμοποιούν δύο ομάδες, οι οποίες τα συμπεριλαμβάνουν είτε στα εικονοστοιχεία υποβάθρου είτε στα εικονοστοιχεία σήματος.



Σχήμα 5.10: Τα διαγράμματα διασποράς και τα διαγράμματα M-A των τριών μεθόδων βασισμένων σε τεχνικές ομαδοποίησης α) Διάγραμμα διασποράς του K-μέσων με 2 χαρακτηριστικά και $K=2$, β) Διάγραμμα M-A του K-μέσων με 2 χαρακτηριστικά και $K=2$ γ) Διάγραμμα διασποράς του K-μέσων με 7 χαρακτηριστικά και $K=2$, δ) Διάγραμμα M-A του K-μέσων με 7 χαρακτηριστικά και $K=2$, ε) Διάγραμμα διασποράς του ασαφούς C-μέσων με 11 χαρακτηριστικά και $C=3$, στ) Διάγραμμα M-A του ασαφούς C-μέσων με 11 χαρακτηριστικά και $C=3$.

Στο σχήμα 5.10 φαίνονται τα διαγράμματα διασποράς (scatter plots) και τα διαγράμματα M-A (*M-A* plots) των τριών προσεγγίσεων. Το διάγραμμα διασποράς είναι το διάγραμμα των λογαριθμικών διορθωμένων τιμών για το κόκκινο κανάλι, έναντι στις λογαριθμικές

διορθωμένες τιμές του πράσινου καναλιού για κάθε μία από της κηλίδες της εικόνας. Το διάγραμμα $M-A$ είναι το διάγραμμα που στον κατακόρυφο άξονα έχει την τιμή M ενώ στον οριζόντιο την τιμή A της κάθε κηλίδας, όπως αυτές ορίστηκαν από της σχέσεις 3.11 και 3.12. Και τα δύο αυτά διαγράμματα παρέχουν μια γενική επισκόπηση του βιολογικού πειράματος.

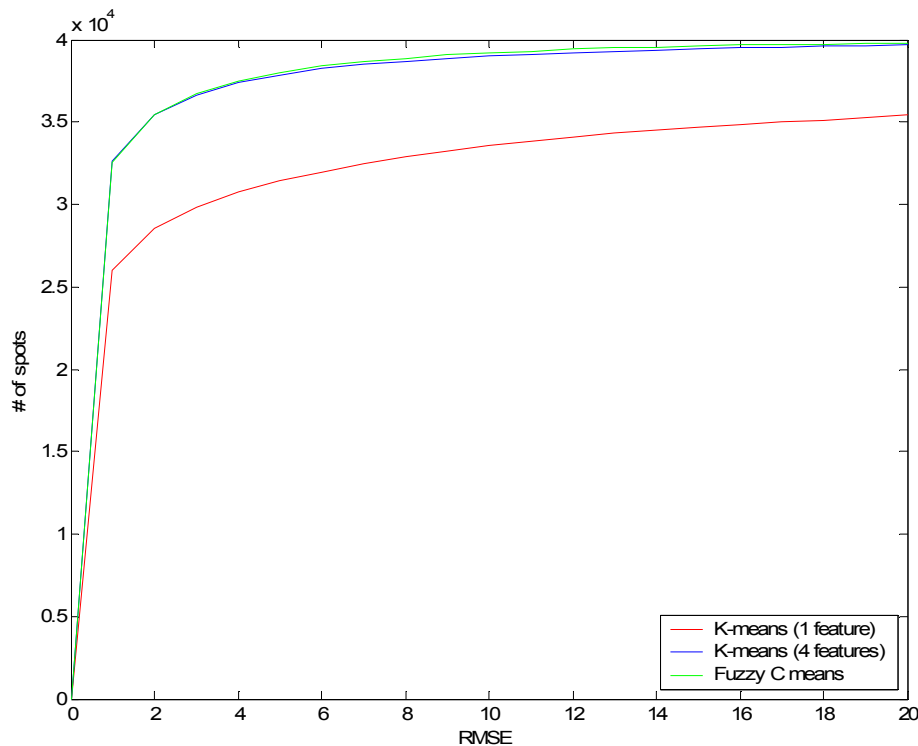
Από την φύση του το πείραμα των μικροσυστοιχιών σκοπεύει στον εντοπισμό ανιχνευτών, οι οποίοι έχουν διαφορετική συμπεριφορά κατά την υβριδοποίηση των δύο δειγμάτων με αυτούς. Το πιθανότερο σε ένα πείραμα μικροσυστοιχιών είναι τα δύο δείγματα να έχουν παρόμοιες υβριδοποιήσεις με τους ανιχνευτές, δηλαδή παρόμοιες ποσοτικοποιημένες τιμές και για τα δύο κανάλια της εικόνας. Οι ανιχνευτές εκείνοι που αποκλίνουν από αυτήν την συμπεριφορά, είναι οι ανιχνευτές ενδιαφέροντος, εφόσον η αλληλουχία τους ενώ βρίσκεται στον ένα από τα δείγματα δεν βρίσκεται στο άλλο. Είναι προφανές για παράδειγμα το ενδιαφέρον που θα παρουσιάζει ένα γονίδιο στις μικροσυστοιχίες cDNA εάν αυτό παρουσιάζει μεγάλο επίπεδο έκφρασης σε ένα καρκινικό δείγμα, ενώ ταυτόχρονα παρουσιάζει μικρό επίπεδο έκφρασης στο αντίστοιχο κανονικό δείγμα. Σύμφωνα με τα παραπάνω αναμένεται η πλειοψηφία των ποσοτικοποιημένων τιμών να είναι παρόμοια στα δύο κανάλια της εικόνας για την ίδια κηλίδα. Αυτό έχει ως αποτέλεσμα τα σημεία στο διάγραμμα διασποράς να συνωστίζονται στην διαγώνιο, ενώ στο διάγραμμα $M-A$ αναμένεται να συνωστίζονται στον οριζόντιο άξονα. Τα σημεία τα οποία δεν ακολουθούν τις παραπάνω συμπεριφορές, είναι αυτά τα οποία προέρχονται από ανιχνευτές που παρουσιάζουν ιδιαίτερο ενδιαφέρον για το εκάστοτε βιολογικό πρόβλημα. Όσον αφορά στην απόδοση των μεθόδων που υλοποιήθηκαν, τα δύο αυτά διάγραμμα μπορούν να δώσουν ένα ποιοτικό συμπέρασμα για την ακρίβεια τους. Όσο μεγαλύτερη είναι η διασπορά των σημείων στα διαγράμματα αυτά, τόσο αναδεικνύεται ότι η μέθοδος δεν υπολογίζει σωστά τις τιμές για την κάθε κηλίδα. Στο σχήμα 5.10 παρατηρούμε ότι οι μέθοδος που χρησιμοποιεί δύο ομάδες και μόνο τις φωτεινότητες των εικονοστοιχείων για χαρακτηριστικά, παρουσιάζει μεγάλη διασπορά των σημείων και στο διάγραμμα διασποράς και στο διάγραμμα $M-A$. Τα σημεία στα διαγράμματα για τις άλλες δύο μεθόδους είναι σαφώς πιο συγκεντρωμένα στην διαγώνιο του διαγράμματος διασποράς και στον οριζόντιο άξονα του διαγράμματος $M-A$.

Με σκοπό να λάβουμε αριθμητικά αποτελέσματα για να αξιολογήσουμε τις μεθόδους, υπολογίστηκε ο λόγος RAT (Εξ. 3.13) μεταξύ των διορθωμένων τιμών για το κόκκινο και το πράσινο κανάλι. Για κάθε μια κηλίδα της εικόνας ο λόγος RAT δίνεται από την επισημείωση της βάσης δεδομένων του Στάνφορντ. Έτσι δίνεται η δυνατότητα της σύγκρισης των επισημειωμένων τιμών του λόγου RAT , με τις τιμές του λόγου που υπολογίστηκαν από τις τρεις προσεγγίσεις. Ένα καλό μέτρο για την σύγκριση είναι το τετραγωνικό σφάλμα (Square Error - SE) των τιμών του λόγου RAT μεταξύ της επισημείωσης, και των προσεγγίσεων που υλοποιήθηκαν. Το μέσο τετραγωνικό σφάλμα δίνεται από την σχέση:

$$SE = \left(RAT^*(k) - \hat{RAT}(k) \right)^2, \quad (5.18)$$

$RAT^*(k)$ είναι ο λόγος RAT της επισημείωσης για την k -οστή κηλίδα και $\hat{RAT}(k)$ είναι ο λόγος RAT που έχει προσδιορίσει η εκάστοτε μέθοδος για την k -οστή κηλίδα της εικόνας.

Για πολλές κηλίδες της εικόνας το τετραγωνικό σφάλμα υπολογίζεται αρκετά υψηλό. Αυτό συμβαίνει κυρίως σε κηλίδες που αντιστοιχούν σε ανιχνευτές που δεν έχουν υβριδοποιηθεί με τα δείγματα. Χρήσιμα συμπεράσματα μπορούν να εξαχθούν δημιουργώντας το διάγραμμα του αριθμού των σωστά καταταμημένων κηλίδων για διάφορες τιμές του τετραγωνικού σφάλματος. Σωστά καταταμημένες κηλίδες για κάθε τιμή του τετραγωνικού σφάλματος, θεωρούμε τις κηλίδες για τις οποίες το τετραγωνικό σφάλμα είναι μικρότερο της τιμής αυτής. Στο σχήμα 5.11 φαίνονται οι καμπύλες, για κάθε μία από τις προσεγγίσεις που υλοποιήσαμε, του αριθμού των κηλίδων που έχουν καταταμηθεί σωστά συναρτήσει των διαφόρων τιμών του τετραγωνικού σφάλματος. Για παράδειγμα, αν θεωρήσουμε τετραγωνικό σφάλμα μικρότερο του δύο ($SE < 2$) τότε παρατηρείται ότι ο αλγόριθμος K-μέσων, με χαρακτηριστικά μόνο τις φωτεινότητες, υπολογίζει σωστά τους λόγους RAT από περίπου 28000 κηλίδες, ενώ οι άλλες δύο υλοποιήσεις προσδιορίζουν σωστά τους λόγους 33000 κηλίδων.



Σχήμα 5.11: Αριθμός σωστά κατατμημένες κηλίδες για κάθε τιμή των τετραγωνικών σφαλμάτων του λόγου RAT.

Στο σχήμα 5.11 παρατηρούμε ότι η πράσινη καμπύλη η οποία αντιστοιχεί στον ασαφή αλγόριθμο C-μέσων με τις τρεις ομάδες, υπερτερεί της μπλε καμπύλης που αντιστοιχεί στον αλγόριθμο K-μέσων για δύο ομάδες. Ωστόσο η διαφορά των δύο υλοποιήσεων δεν είναι πολύ μεγάλη. Για κάποιες τιμές του τετραγωνικού σφάλματος φτάνει περίπου στις 500 περισσότερες σωστά κατατμημένες κηλίδες. Είναι σημαντικό να αναφέρουμε βέβαια ότι ο τρόπος με τον οποίο έχουν επισημειωθεί οι εικόνες της βάσης του Στάνφορντ (με την χρήση του προγράμματος ScanAlyze) προσεγγίζεται πιο καλά από την χρήση δύο ομάδων και αυτό έχει αντίκτυπο στα αποτελέσματα. Όπου υπάρχουν σκοτεινές οπές η επισημείωση έχει συμπεριλάβει τα εικονοστοιχεία τους ως εικονοστοιχεία σήματος, ενώ τα εικονοστοιχεία από τεχνουργήματα έχουν συμπεριληφθεί πάντα στα εικονοστοιχεία υποβάθρου. Αυτό σημαίνει ότι μια προσέγγιση δύο ομάδων συμφωνεί περισσότερο με την επισημείωση των συγκεκριμένων εικόνων. Παρά το γεγονός αυτό όμως ο ασαφής αλγόριθμος C-μέσων με ολόκληρο το σύνολο των χαρακτηριστικών παρουσιάζει καλύτερα αποτελέσματα από τις άλλες δύο προσεγγίσεις.

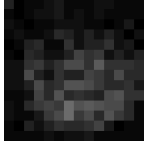

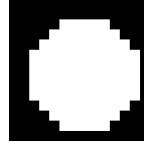
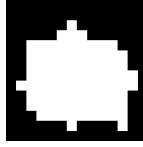
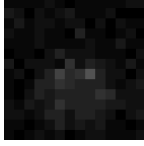

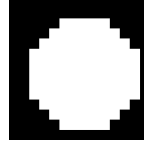
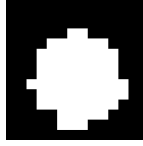


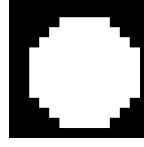

Στο τελευταίο εδάφιο του κεφαλαίου παρουσιάζονται αποτελέσματα της ακρίβειας για τις παραπάνω προσεγγίσεις σε επίπεδο εικονοστοιχείου, με σκοπό να συγκριθούν με τις προσεγγίσεις που υλοποιήθηκαν με κατάτμηση με επίβλεψη.

5.6.2.Κατάτμηση με επίβλεψη

Η υλοποίηση της κατάτμησης με επίβλεψη ανέδειξε και την αναγκαιότητα για αποτελέσματα σε επίπεδο εικονοστοιχείου για την αξιολόγηση των μεθόδων. Όπως αναφέρθηκε στο εδάφιο 5.5, με κατάλληλες παρεμβάσεις εξάγαμε την επισημείωση σε επίπεδο εικονοστοιχείων για το σύνολο των εικόνων από την βάση δεδομένων του Στάνφορντ, αλλά και για τις εικόνες που προέκυψαν από την προσομοίωση. Έτσι στάθηκε δυνατός ο υπολογισμός της ακρίβειας των μεθόδων σε επίπεδο εικονοστοιχείων. Η ακρίβεια αυτή εκφράζει το ποσοστό επιτυχίας του χαρακτηρισμού των εικονοστοιχείων σύμφωνα με τον χαρακτηρισμό της επισημείωσης. Παρακάτω θα παρουσιαστούν τα αποτελέσματα της κατάτμησης με επίβλεψη, αρχικά για τις πραγματικές εικόνες από την βάση δεδομένων του Στάνφορντ, για τις οποίες η επισημείωση χαρακτηρίζει τα εικονοστοιχεία σε δύο κατηγορίες. Εν συνεχεία θα παρουσιαστούν τα αποτελέσματα για τις προσομοιωμένες εικόνες για τις οποίες υπάρχει στην επισημείωση και ο χαρακτηρισμός για τα εικονοστοιχεία τεχνουργημάτων.

5.6.2.1.Πραγματικές εικόνες

Στο σχήμα 5.12 φαίνεται το αποτέλεσμα της εφαρμογής της κατάτμησης με επίβλεψη για τρεις κηλίδες της εικόνας. Συγκεκριμένα το αποτέλεσμα αυτό προέκυψε από την εφαρμογή των Μηχανών Διανυσμάτων Υποστήριξης, όπως αυτά παρουσιάστηκαν στην εργασία των Giannakeas *et al.* (2010) [189]. Στο σχήμα αυτό έχουν επιλεγεί μια κηλίδα πολύ υψηλής φωτεινότητας, μια κηλίδα με κανονικά επίπεδα φωτεινότητας και μια κηλίδα χαμηλής φωτεινότητας.

| Περιγραφή | Κόκκινο Κανάλι | Πράσινο Κανάλι | Επισημείωση | Αποτέλεσμα |
|-----------------------------|---|---|--|---|
| Υβριδοποιημένη κηλίδα |  |  |  |  |
| Κηλίδα χαμηλής Φωτεινότητας |  |  |  |  |
| Κηλίδα υψηλής Φωτεινότητας |  |  |  |  |

Σχήμα 5.12: Οπτικοποίηση των αποτελεσμάτων της κατάτμησης με επίβλεψη.

Για να αξιολογήσουμε την μέθοδο για το πρόβλημα δύο κλάσεων των πραγματικών εικόνων ορίσαμε την ευαισθησία (Se), την εξειδίκευση (Sp) και την ακρίβεια (Acc) ως εξής:

$$Se = \frac{\#of\ correctly\ identified\ background\ pixels}{total\ \#of\ background\ pixels}, \quad (5.19)$$

$$Sp = \frac{\#of\ correctly\ identified\ signal\ pixels}{total\ \#of\ signal\ pixels}, \quad (5.20)$$

$$Acc = \frac{\#of\ correctly\ identified\ pixels}{total\ \#of\ pixels}. \quad (5.21)$$

Στον πίνακα 5.2 παρουσιάζονται τα αποτελέσματα της ευαισθησίας, της εξειδίκευσης και της ακρίβειας για τις πραγματικές εικόνες από την βάση δεδομένων του Στάνφορντ, με χρήση του Μπεϋζιανού ταξινομητή και των Μηχανών Διανυσμάτων Υποστήριξης.

Πίνακας 5.2: Τα αποτελέσματα της ευαισθησίας, της εξειδίκευσης και της ακρίβειας για τις πραγματικές εικόνες με χρήση του Μπεϋζιανού ταξινομητή και των Μηχανών Διανυσμάτων Υποστήριξης.

| Ταξινομητής | <i>Sp</i> (%) | <i>Se</i> (%) | <i>Acc</i> (%) |
|--------------------|-------------------------|-------------------------|--------------------------|
| Bayes | 83.25 | 89.19 | 86.22 |
| SVM | 96.62 | 98.59 | 97.59 |

Όπως φαίνεται στον πίνακα 5.2 η κατάτμηση με επίβλεψη παρουσιάζει υψηλά ποσοστά ακρίβειας. Συγκεκριμένα η ακρίβεια της μεθόδου με χρήση Μπεϋζιανού ταξινομητή υπολογίζεται σε 86.22% ενώ με χρήση Μηχανών Διανυσμάτων Υποστήριξης ξεπερνά το 97%. Το γεγονός ότι οι Μηχανές Διανυσμάτων Υποστήριξης είναι ένας πιο εξελιγμένος αλγόριθμος σε σχέση με τον απλό Μπεϋζιανό ταξινομητή δικαιολογεί μια τέτοια διαφορά. Μια καλύτερη εικόνα της αξιοπιστίας της κατάτμησης με επίβλεψη θα διαφανεί συγκρίνοντας την προσέγγιση αυτή με προσεγγίσεις οι οποίες στηρίζονται σε τεχνικές ομαδοποίησης, όπως θα δούμε αργότερα.

5.6.2.2. Προσομοιωμένες εικόνες

Με παρόμοιο τρόπο με αυτόν που παρουσιάστηκαν τα αποτελέσματα της κατάτμησης με επίβλεψη για τις πραγματικές εικόνες από την βάση δεδομένων του Στάνφορντ, για την οπτικοποίηση των αποτελεσμάτων επιλέχθηκαν τρεις ενδεικτικές κηλίδες από κάθε προσομοιωμένο σύνολο εικόνων (καλές, κανονικές και κακές εικόνες). Τα αποτελέσματα αυτά προέκυψαν με την χρήση Μηχανών Διανυσμάτων Υποστήριξης. Στο σχήμα 5.13 απεικονίζονται σε κάθε στήλη το κόκκινο κανάλι, το πράσινο κανάλι, η εικόνα επισημείωσης, και το αποτέλεσμα της κατάτμησης με επίβλεψη αντιστοίχως. Στις εικόνες της επισημείωσης και στις εικόνες του αποτελέσματος τα εικονοστοιχεία σήματος επισημαίνονται με λευκό χρώμα, τα εικονοστοιχεία υποβάθρου με μαύρο ενώ τα εικονοστοιχεία της τρίτης κατηγορίας των τεχνουργημάτων με γκρι χρώμα.

| | | Κόκκινο Κανάλι | Πράσινο Κανάλι | Επισημ. | Αποτέλ. |
|----------------------|--|---------------------------|---------------------------|----------------|----------------|
| Καλές Εικόνες | Υβριδοποιημένη Κηλίδα | | | | |
| | Κηλίδα χαμηλής Φωτεινότητας | | | | |
| | Κηλίδα υψηλής φωτεινότητας | | | | |
| Κανονικές Εικόνες | Κηλίδα χαμηλής Φωτεινότητας | | | | |
| | Κηλίδα υψηλής φωτεινότητας | | | | |
| | Κηλίδα με οπή (donut) | | | | |
| Κακές Εικόνες | Κηλίδα με θόρυβο υποβάθρου | | | | |
| | Κηλίδα με τεχνούργημα | | | | |
| | Κηλίδα χαμηλής φωτεινότητας με τεχνούργημα | | | | |

Σχήμα 5.13: Οπτικοποίηση των αποτελεσμάτων της κατάτμησης με επίβλεψη.

Όπως φαίνεται στο σχήμα 5.13, η μέθοδος εντοπίζει με ακρίβεια τις κηλίδες υψηλής φωτεινότητας. Στον αντίποδα υπάρχουν περιπτώσεις όπου η κηλίδα δεν είναι ξεκάθαρη στην εικόνα με αποτέλεσμα η μέθοδος να δυσκολεύεται να χαρακτηρίσει σωστά τα εικονοστοιχεία. Τέτοιες περιπτώσεις συμβαίνουν όταν η κηλίδα είναι χαμηλής φωτεινότητας και υπάρχει θόρυβος, είτε υπό την μορφή θορύβου υποβάθρου, είτε υπό την μορφή τεχνουργημάτων. Όπως φαίνεται στο σχήμα 5.13 σε κάποιες από τις περιπτώσεις έχουν χαρακτηριστεί λάθος τα εικονοστοιχεία τεχνουργημάτων. Ωστόσο, τα εικονοστοιχεία αυτά δεν θα συμπεριληφθούν στην ποσοτικοποίηση της υβριδοποίησης της κάθε κηλίδας και αυτό είναι και το νόημα της εισαγωγής αυτής της κατηγορίας. Τα εικονοστοιχεία τα οποία είναι αμφισβητήσιμο αν πρέπει να συμμετέχουν στην ποσοτικοποίηση ως εικονοστοιχεία σήματος ή υποβάθρου δεν πρέπει να λαμβάνονται υπόψη.

Σε αντιστοιχία με το πρόβλημα των δύο κλάσεων για να εξαχθούν αριθμητικά αποτελέσματα για την αξιοπιστία της μεθόδου, ορίσαμε την ακρίβεια μέσω του πίνακα σύγχυσης (Confusion Matrix). Όσο ο αλγόριθμος χαρακτηρίζει τα εικονοστοιχεία ο πίνακας σύγχυσης υπολογίζεται από την σχέση:

$$CM(\bar{c}_i, c_i^*) = CM(\bar{c}_i, c_i^*) + 1, \quad (5.22)$$

όπου \bar{c}_i είναι η πραγματική κλάση που ανήκει το εικονοστοιχείο, και c_i^* η κλάση στην οποία το κατατάσσει η μέθοδος. Η ακρίβεια της μεθόδου δίνεται από τον τύπο:

$$Acc = \frac{trace(CM)}{\# \text{ of all pixels}}. \quad (5.22)$$

Στον πίνακα 5.3 παρουσιάζονται οι τιμές της ακρίβειας της μεθόδου για το πρόβλημα τριών κλάσεων των προσομοιωμένων εικόνων.

Πίνακας 5.3: Ακρίβεια της μεθόδου για το πρόβλημα τριών κλάσεων των προσομοιωμένων εικόνων.

| | Σύνολο Εκπαίδευσης | Σύνολο Ελέγχου | Dataset | | |
|--------------|-----------------------|-------------------|---------|--------|-------|
| | | | Good | Normal | Bad |
| Bayes | 50 spots | 300 spots | 95.52 | 71.24 | 55.18 |
| SVM | 14450 pixels | 86700 pixels | 96.03 | 93.24 | 71.35 |

Όπως φαίνεται στο σχήμα η εφαρμογή των Μηχανών Διανυσμάτων Υποστήριξης παρουσιάζει ακρίβεια μεγαλύτερη από 93% για τις καλές και τις κανονικές εικόνες. Αντίθετα η ακρίβεια πέφτει στο 71% για τις κακές προσομοιωμένες εικόνες οι οποίες είναι έχουν πολύ θόρυβο. Στον πίνακα 5.3 φαίνεται επίσης η αξιοπιστία των Μηχανών Διανυσμάτων Υποστήριξης σε σχέση με τον Μπεϋζιανό ταξινομητή. Ο Μπεϋζιανός ταξινομητής φαίνεται να παρουσιάζει παραπλήσια ακρίβεια με της Μηχανές Διανυσμάτων Υποστήριξης για τις καλές εικόνες, όσο όμως πέφτει η ποιότητα των εικόνων αδυνατεί να ανταποκριθεί.

5.7. Συμπεράσματα

Στο κεφάλαιο αυτό αναπτύχθηκε η μεθοδολογία για την κατάτμηση εικόνων μικροσυστοιχιών. Η κατάτμηση ξεκινά με την εξαγωγή των χαρακτηριστικών για κάθε εικονοστοιχείο του κελιού μιας κηλίδας. Η μέθοδος προτείνει ένα καινοτόμο διάλυμα χαρακτηριστικών, το οποίο στοχεύει στον καλύτερο διαχωρισμό των εικονοστοιχείων σήματος και υποβάθρου αλλά και στον εντοπισμό των εικονοστοιχείων τεχνουργημάτων. Την κατηγορία αυτήν των τεχνουργημάτων την εισάγαμε στο πρόβλημα με σκοπό να καταχωρηθούν σε αυτή τα εικονοστοιχεία τα οποία πρέπει να εξαιρεθούν από την διαδικασία της ποσοτικοποίησης. Δεδομένου ότι κάποια από τα χαρακτηριστικά δεν είναι ανεξάρτητα μεταξύ τους, εφαρμόζεται η Ανάλυση Κυρίων Συνιστωσών, για την μείωση της διάστασης του διανύσματος χαρακτηριστικών και την εξαγωγή ανεξάρτητων γραμμικών συνδυασμών τους. Για τον διαχωρισμό των εικονοστοιχείων της εικόνας εφαρμόστηκαν δύο ειδών μέθοδοι βασισμένες σε τεχνικές μηχανικής μάθησης. Ο αλγόριθμος K-μέσων και ο ασαφής αλγόριθμος C-μέσων χρησιμοποιήθηκαν για την προσέγγιση που είναι βασισμένη σε τεχνικές ομαδοποίησης, ενώ ο Μπεϋζιανός ταξινομητής και ο ταξινομητής μηχανών διανυσμάτων υποστήριξης εφαρμόστηκαν για την προσέγγιση με επίβλεψη που είναι βασισμένη σε τεχνικές ταξινόμησης. Είναι η πρώτη

φορά που εφαρμόζονται τεχνικές ταξινόμησης στο πεδίο της κατάτμησης των μικροσυστοιχιών.

Πίνακας 5.4: Συγκριτικά αποτελέσματα μεθόδων χωρίς επίβλεψη και μεθόδων με επίβλεψη.

| Εργασία | Αλγόριθμος | Χαρακτηριστικά | Σύνολο δεδομένων (Acc %) | | | |
|---------------------------------------|------------|--|--------------------------|------------|--------------|-----------|
| | | | Πραγ. | Προσ. Good | Προσ. Normal | Προσ. Bad |
| Bozinov <i>et al.</i> [154] | K-means | $I_R(i, j), I_G(i, j)$ | 61.9 | 64.9 | 61.6 | 44.6 |
| Ergüt <i>et al.</i> [155] | FCM | $I_R(i, j), I_G(i, j)$ | 62.4 | 65.0 | 61.5 | 43.6 |
| Wu <i>et al.</i> [156] | K-means | $I_R(i, j), I_G(i, j), i, j, D_{Eucl}$ | 70.3 | 64.7 | 63.4 | 44.8 |
| - | K-means | $I_R(i, j), I_G(i, j), \bar{I}_R^{Window}, \bar{I}_G^{Window}, \sigma_R, \sigma_G, i, j, D_{Eucl}, \text{Χαρακτηριστικά σχήματος}$ | 69.1 | 73.0 | 67.3 | 50.5 |
| Giannakeas <i>et al.</i> , 2009 [180] | FCM | $I_R(i, j), I_G(i, j), \bar{I}_R^{Window}, \bar{I}_G^{Window}, \sigma_R, \sigma_G, i, j, D_{Eucl}, \text{Χαρακτηριστικά σχήματος}$ | 69.4 | 78.4 | 67.8 | 49.0 |
| - | K-means | Όλα τα χαρακτηριστικά | 70.5 | 73.5 | 70.1 | 50.9 |
| - | FCM | Όλα τα χαρακτηριστικά | 71.7 | 78.3 | 69.7 | 49.4 |
| Giannakeas <i>et al.</i> 2007 [87] | Bayes | Όλα τα χαρακτηριστικά | 86.2 | 95.5 | 71.2 | 55.2 |
| Giannakeas <i>et al.</i> 2010 [188] | SVM | Όλα τα χαρακτηριστικά | 97.6 | 96.0 | 93.2 | 71.3 |

Στο παρόν εδάφιο επιχειρείται μια άμεση σύγκριση των δύο προσεγγίσεων, υλοποιώντας παραλλαγές των δύο προσεγγίσεων, οι οποίες έχουν παρουσιαστεί στο παρελθόν. Η σύγκριση αυτή αναδεικνύει τα πλεονεκτήματα των προσεγγίσεων με επίβλεψη έναντι των προσεγγίσεων χωρίς επίβλεψη. Συγκεκριμένα, για να συγκρίνουμε τις δύο προσεγγίσεις χρησιμοποιήσαμε τους αλγορίθμους και τα σύνολα χαρακτηριστικών που αναφέρθηκαν ανωτέρω όπως αυτά παρουσιάστηκαν στις εργασίες [180, 187-188]. Επίσης χρησιμοποιήσαμε για τις μεθόδους χωρίς επίβλεψης το σύνολο των χαρακτηριστικών, για το οποίο δεν έχει παρουσιαστεί εργασία στην βιβλιογραφία. Αυτό έγινε για να υπάρχει

άμεση σύγκριση των δύο προσεγγίσεων με το ίδιο σύνολο χαρακτηριστικών. Για την εξαγωγή της ακρίβειας της κάθε προσέγγισης χρησιμοποιήθηκαν τα 4 διαφορετικά σύνολα δεδομένων (πραγματικές εικόνες, good – normal – bad προσομοιωμένες εικόνες) προσδιορίζοντας την ακρίβεια σε επίπεδο εικονοστοιχείου. Τα αποτελέσματα παρουσιάζονται στον πίνακα 5.4.

Όσον αφορά στην συνεισφορά του διανύσματος χαρακτηριστικών από τον πίνακα 5.4 συμπεραίνουμε ότι όσο προστίθενται περισσότερα χαρακτηριστικά η ακρίβεια της μεθόδου αυξάνεται. Συγκεκριμένα η προσέγγιση των Wu *et al.* [156], η οποία χρησιμοποιεί και χωρικά χαρακτηριστικά υπερτερεί της μεθόδου των Bozinov *et al.* [154] οι οποίοι χρησιμοποιούν μόνο την φωτεινότητα των εικονοστοιχείων. Αντίστοιχα παρόμοια αύξηση στην ακρίβεια της μεθόδου παρατηρείται συγκρίνοντας την εργασία των Wu *et al.* και τις προσεγγίσεις που χρησιμοποιούν ολόκληρο το προτεινόμενο διάνυσμα χαρακτηριστικών.

Το βασικό συμπέρασμα που εξάγεται από τον πίνακα 5.4 είναι η μεγάλη διαφορά που παρουσιάζεται στην ακρίβεια μεταξύ των προσεγγίσεων της κατάτμησης με ή χωρίς επίβλεψη. Όπως φαίνεται στον πίνακα 5.4, στις περισσότερες περιπτώσεις η μεγαλύτερη ακρίβεια των προσεγγίσεων με επίβλεψη υπερβαίνει άνω των είκοσι ποσοστιαίων μονάδων την μεγαλύτερη ακρίβεια των προσεγγίσεων χωρίς επίβλεψη. Η αξιοσημείωτη αυτή διαφορά σε ένα ποσοστό οφείλεται στο γεγονός ότι οι ταξινομητές εκπαιδεύονται και είναι σαφώς πιο αποδοτικοί από τις τεχνικές ομαδοποίησης. Ωστόσο, ο λόγος αυτός δεν είναι ο κυριότερος λόγος που προκαλεί την μεγάλη διαφορά στα ποσοστά της ακρίβειας. Η κατάτμηση που είναι βασισμένη στους ταξινομητές χαρακτηρίζει άμεσα το κάθε ένα εικονοστοιχείο σε μία κατηγορία. Αντιθέτως, η κατάτμηση που είναι βασισμένη σε τεχνικές ομαδοποίησης πρώτα δημιουργεί ομάδες εικονοστοιχείων και εν συνεχεία χαρακτηρίζει την κάθε ομάδα. Έτσι τα εικονοστοιχεία χαρακτηρίζονται έμμεσα από τον χαρακτηρισμό της ομάδας. Σύμφωνα με τον κανόνα της υψηλότερης φωτεινότητας η ομάδες με υψηλή φωτεινότητα χαρακτηρίζονται ως σήμα, ενώ η ομάδα με την χαμηλή μέση φωτεινότητα χαρακτηρίζεται ως υπόβαθρο. Αυτός ο κανόνας πολύ δύσκολα θα κάνει λάθος στα προβλήματα δύο κλάσεων, διότι είναι σχεδόν απίθανο μέσα σε ένα κελί μια

κηλίδας το υπόβαθρο να έχει υψηλότερη φωτεινότητα από το σήμα, ακόμα και στις περιπτώσεις των μη-υβριδοποιημένων κηλίδων. Στην περίπτωση όμως του προβλήματος των τριών κλάσεων, όπου υφίσταται και η κατηγορία των τεχνουργημάτων, ο κανόνας αυτός ενδέχεται να αποτύχει. Η αποτυχία του κανόνα συμβαίνει κυρίως όταν υπάρχουν φωτεινά τεχνουργήματα στο κελί της κηλίδας. Σε αυτήν την περίπτωση η κατηγορία των τεχνουργημάτων θα χαρακτηριστεί λανθασμένα ως σήμα. Οι επιπτώσεις στην ακρίβεια της μεθόδου σε επίπεδο εικονοστοιχείων είναι πολύ μεγάλες. Αμέσως χαρακτηρίζονται λανθασμένα όλα τα εικονοστοιχεία σήματος και όλα τα εικονοστοιχεία τεχνουργημάτων, το μεγαλύτερο μέρος δηλαδή των εικονοστοιχείων του κελίου. Αυτή είναι και η κυριότερη αιτία όπου η κατάτμηση χωρίς επίβλεψη έχει τόσο μεγάλη διαφορά σε ποσοστά ακρίβεια σε σχέση με την κατάτμηση με επίβλεψη. Με τον άμεσο χαρακτηρισμό του κάθε εικονοστοιχείου που γίνεται από τους ταξινομητές αποφεύγεται ο μαζικός λανθασμένος χαρακτηρισμός εικονοστοιχείων.

Κεφάλαιο 6^ο: Συμπεράσματα

6.Συμπεράσματα

Οι μικροσυστοιχίες αποτελούν μια τεχνολογία η οποία έχει αναπτυχθεί για βιολογικούς και γενετικούς σκοπούς τα τελευταία δέκα πέντε περίπου χρόνια. Η ανάπτυξη των μικροσυστοιχιών αυτά τα χρόνια είναι ραγδαία, ενώ ανάλογες είναι και οι απαιτήσεις της τεχνολογίας, από τα πληροφοριακά συστήματα που θα κλιθούν να επεξεργαστούν τον μεγάλο όγκο των δεδομένων που παράγονται, έως τον ίδιο το εξοπλισμό για την διεξαγωγή του πειράματος. Η παρέμβαση της επιστήμης της πληροφορικής στις μικροσυστοιχίες γίνεται σε δύο διαφορετικά επίπεδα. Ξεκινάει από την επεξεργασία των εικόνων που παράγονται από την σάρωση της μικροσυστοιχίας. Σκοπός της επεξεργασίας της εικόνας είναι η εξαγωγή των ποσοτικοποιημένων τιμών της υβριδοποίησης που λαμβάνει χώρα κατά την διάρκεια του βιολογικού πειράματος. Εν συνεχεία η ποσοτικοποιημένες τιμές της υβριδοποίησης των δειγμάτων με τον κάθε ανιχνευτή επεξεργάζονται περαιτέρω με σκοπό την εξαγωγή νέας βιολογικής γνώσης.

Στην παρούσα διδακτορική διατριβή αναπτύχθηκαν ευφυείς μέθοδοι για την επεξεργασία εικόνων μικροσυστοιχιών. Η συνεισφορά της διατριβής στην παρούσα στάθμιση της επεξεργασίας των εικόνων των μικροσυστοιχιών αναδεικνύεται στα δύο βασικά στάδια

της. Τα καινοτόμα στοιχεία υλοποιήθηκαν άμεσα στο στάδιο του εντοπισμού των κηλίδων της εικόνας, αλλά και στο στάδιο της κατάτμησης των κηλίδων. Έμμεσος είναι ο αντίκτυπος τους στην εξαγωγή των εντάσεων της κάθε κηλίδας. Συνοπτικά, κατά την διάρκεια της διατριβής ερευνήθηκε το πεδίο του εντοπισμού των κηλίδων στην εικόνα, καταλήγοντας σε μία γενικευμένη μέθοδο, η οποία καλύπτει όλο το φάσμα των εικόνων που εξάγονται από τις υπάρχουσες τεχνολογίες. Σκοπός της μεθόδου ήταν να αντιμετωπιστεί και ο ολοένα αυξανόμενος αριθμός των ανιχνευτών, και κατ' επέκταση των κηλίδων στην εικόνα. Σχετικά με το στάδιο της κατάτμησης η παρούσα διατριβή συμβάλει στον ορθολογικό διαχωρισμό των εικονοστοιχείων σήματος από τα εικονοστοιχεία υποβάθρου. Αυτό επιτυγχάνεται κυρίως με την χρήση τεχνικών ταξινόμησης και με την επέκταση του διανύσματος χαρακτηριστικών για κάθε ένα εικονοστοιχείο. Επίσης συμβάλει σε αυτό η ιδέα του αποκλεισμού των εικονοστοιχείων, η φωτεινότητα των οποίων έχει αμφισβητούμενη προέλευση. Τα παραπάνω δίδουν την δυνατότητα στην περισσότερο αξιόπιστη εξαγωγή των εντάσεων για κάθε μια υβριδοποίηση των ανιχνευτών με τα δείγματα.

Για τον εντοπισμό των κηλίδων ακολουθήσαμε μια διαδικασία τεσσάρων βημάτων, η οποία συνεχώς εξελίσσονταν καταλήγοντας σε μια γενικευμένη μέθοδο για την επεξεργασία εικόνων που προέρχονται είτε από μικροσυστοιχίες με τετραγωνική χωροθέτηση των ανιχνευτών είτε από μικροσυστοιχίες με εξαγωνική χωροθέτηση. Με τον τρόπο αυτό η μέθοδος που αναπτύχθηκε εξυπηρετεί την ποικιλομορφία των μικροσυστοιχιών που έχει προκύψει από τις διαφορετικές προσεγγίσεις της εκτύπωσης των ανιχνευτών σε μια μικροσυστοιχία. Και στα τέσσερα βήματα που ακολουθήθηκαν προστέθηκαν καινοτόμα σημεία με σκοπό τον πιο αξιόπιστο εντοπισμό των κηλίδων.

Περισσότερο σημαντική είναι η συνεισφορά του αλγορίθμου των αναπτυσσόμενων ομόκεντρων πολυγώνων για τον εντοπισμό των μη υβριδοποιημένων κηλίδων. Ο αλγόριθμος αυτός προχώρησε ένα βήμα παραπέρα τις προσεγγίσεις κηλίδα με κηλίδα, εκμεταλλευόμενος τις κοινές ιδιότητες των δύο πλεγμάτων. Όπως αναπτύξαμε στο 4^ο κεφάλαιο η προσέγγιση αυτή εντοπίζει σε κάθε επανάληψη την θέση μίας κηλίδας λαμβάνοντας υπόψη μόνο την θέση γειτονικών φωτεινών κηλίδων. Αυτόν τον περιορισμό

είχε και η αρχική προσέγγιση που υλοποιήσαμε στα πλαίσια αυτής της διατριβής για τον εντοπισμό κηλίδων σε εικόνες τετραγωνικού πλέγματος, γεγονός το οποίο την άφηνε εκτεθειμένη σε τοπικού χαρακτήρα προβλήματα της εικόνας. Με την γενικευμένη μέθοδο, εκμεταλλευόμενοι τις κοινές ιδιότητες των πλεγμάτων προσδιορίζονται πολλές θέσεις των σκοτεινών κηλίδων σε κάθε επανάληψη του αλγορίθμου, καθιστώντας εφικτή την παραλληλοποίηση της διαδικασίας. Ο δε προσδιορισμός των θέσεων δεν γίνεται τοπικά μόνο από τις θέσεις των γειτονικών κηλίδων αλλά προσδιορίζονται σύμφωνα με τις ιδιότητες ολόκληρου του πλέγματος από κηλίδες.

Ένα άλλο πλεονέκτημα της μεθοδολογίας απορρέει από το γεγονός ότι κατά την εκτέλεση του αλγορίθμου των αναπτυσσόμενων πολυγώνων το πλέγμα με τα κέντρα των κηλίδων ξαναχτίζεται από την αρχή, και επαληθεύεται με τα κέντρα που έχουν βρεθεί από το βήμα του εντοπισμού των φωτεινών κηλίδων. Κατ' αυτόν τον τρόπο από τα εντοπισμένα φωτεινά αντικείμενα που έχουν βρεθεί στο πλέγμα θα παραμείνουν μόνο αυτό που αντιστοιχούν σε πραγματικές κηλίδες, απαλλάσσοντας το πλέγμα από κέντρα τεχνουργημάτων.

Όσον αφορά στο στάδιο της κατάτμησης των εικόνων η παρούσα διατριβή προσέφερε τρεις βασικές καινοτομίες ώστε να βελτιωθεί ο χαρακτηρισμός των εικονοστοιχείων. Η σωστή επιλογή των εικονοστοιχείων που θα χρησιμοποιηθούν για να εξαχθούν οι τιμές της υβριδοποίησης των ανιχνευτών με τα δείγματα, είναι το πιο σημαντικό σημείο για το στάδιο της επεξεργασίας των μικροσυστοιχιών. Για την σωστή επιλογή των εικονοστοιχείων εφαρμόστηκαν τεχνικές ταξινόμησης για την κατάτμηση της εικόνας, χρησιμοποιήθηκε τρίτη κλάση εικονοστοιχείων στην οποία και δόθηκε φυσική υπόσταση, ενώ τέλος επεκτάθηκε το διάνυμα των χαρακτηριστικών.

Για την πιο αξιόπιστη κατάτμηση των εικονοστοιχείων της εικόνας με χρήση τεχνικών μηχανικής μάθησης, η παρούσα διατριβή επέκτεινε το διάνυμα των χαρακτηριστικών των εικονοστοιχείων. Όπως αναφέρθηκε οι μέθοδοι που έχουν παρουσιαστεί στην βιβλιογραφία τροφοδοτούν τους αλγορίθμους μόνο με την φωτεινότητα των εικονοστοιχείων. Δεδομένης της ύπαρξης τεχνουργημάτων στις εικόνες μικροσυστοιχιών

μόνο η φωτεινότητα των εικονοστοιχείων δεν αρκεί για να χαρακτηρίσει την φύση των εικονοστοιχείων. Για τον λόγο αυτό στο διάνυσμα χαρακτηριστικών προστέθηκαν χαρακτηριστικά που αφορούν την φωτεινότητα της γειτονιάς των εικονοστοιχείων, γεωμετρικά χαρακτηριστικά και τέλος χαρακτηριστικά που αναδεικνύουν την ομοιότητα της γειτονιάς του εικονοστοιχείου με την σχηματομορφή μιας τυπικής κηλίδας στην εικόνα. Τα αποτελέσματα που παρουσιάστηκαν, αποδεικνύουν ότι η προσθήκη των επιπλέον χαρακτηριστικών βελτίωσαν κατά πολύ την απόδοση της μεθόδου σε συνδυασμό με την εισαγωγή της τρίτης κατηγορίας εικονοστοιχείων.

Η τρίτη κατηγορία εικονοστοιχείων εισήχθη με σκοπό να κατηγοριοποιήσουμε τα εικονοστοιχεία, οι φωτεινότητες των οποίων δεν θα πρέπει να συμπεριληφθούν στους υπολογισμούς των ποσοτικοποιημένων τιμών. Τα εικονοστοιχεία τεχνουργημάτων είναι εικονοστοιχεία υψηλής φωτεινότητας, τα οποία πιθανότατα θα συμπεριλαμβάνονταν στα εικονοστοιχεία σήματος, αλλοιώνοντας έτσι τις ποσοτικοποιημένες τιμές. Επίσης από την ποσοτικοποίηση εξαιρούνται τα εικονοστοιχεία των σκοτεινών οπών που πολλές φορές βρίσκονται μέσα στις κηλίδες με σχήμα ντόνατς και τα εικονοστοιχεία από το περίγραμμα της κηλίδας. Τα τρία είδη εικονοστοιχείων που έχει αυτή η κατηγορία συνήθως έχουν παρόμοιες φωτεινότητες και αυτός είναι ο λόγος που τα συμπεριλάβαμε σε μια ενιαία κατηγορία. Με τις υπάρχουσες μεθόδους τα εικονοστοιχεία αυτά λαμβάνονταν είτε στον υπολογισμό των εικονοστοιχείων σήματος είτε στον υπολογισμό των εικονοστοιχείων υποβάθρου κατά περίπτωση. Και στις δύο περιπτώσεις οι τελικές τιμές υβριδοποίησης υπολογίζονταν αλλοιωμένες. Αποκλείοντας όμως τα εικονοστοιχεία αυτά από τους υπολογισμούς δεν υπάρχει κανένα αντίκτυπο στις τιμές της υβριδοποίησης.

Ίσως το πιο σημαντικό από τα αποτελέσματα της διατριβής εξήχθη από την σύγκριση των μεθόδων για την κατάτμηση που είναι βασισμένες σε τεχνικές μηχανικής μάθησης (Πίνακας 5.4). Από αυτό διαπιστώνει κανείς ότι η χρήση των ταξινομητών στο πεδίο της κατάτμησης των μικροσυστοιχείων πρέπει να επεκταθεί περαιτέρω. Η αξιοσημείωτη διαφορά των μεθόδων σε ένα ποσοστό οφείλεται στο γεγονός ότι οι ταξινομητές εκπαιδεύονται και είναι σαφώς πιο αποδοτικοί από τις τεχνικές ομαδοποίησης. Η κατάτμηση που είναι βασισμένη στους ταξινομητές πλεονεκτεί όμως επίσης διότι

χαρακτηρίζει άμεσα το κάθε ένα εικονοστοιχείο σε μία κατηγορία. Αντιθέτως, η κατάτμηση που είναι βασισμένη σε τεχνικές ομαδοποίησης πρώτα δημιουργεί ομάδες εικονοστοιχείων και εν συνεχεία χαρακτηρίζει την κάθε ομάδα. Έτσι τα εικονοστοιχεία χαρακτηρίζονται έμμεσα από τον χαρακτηρισμό της ομάδας. Η χρήση λοιπόν τεχνικών ταξινόμησης που προτάθηκε από την παρούσα διδακτορική διατριβή αποδεικνύεται πολύ αποτελεσματική για τον αξιόπιστο διαχωρισμό των εικονοστοιχείων της εικόνας.

Όσον αφορά στην εξαγωγή των ποσοτικοποιημένων τιμών της υβριδοποίησης οι παρεμβάσεις που μπορεί να κάνεις δεν είναι πολλές. Οι τιμές οι οποίες εξάγονται για κάθε κηλίδα είναι συγκεκριμένες και αποσκοπούν στην σύγκριση της υβριδοποίησης των ανιχνευτών για τα διαφορετικά δείγματα. Ωστόσο έμμεση είναι η συνεισφορά σε αυτό το πεδίο όσο πιο αξιόπιστη είναι η κατάτμηση των εικονοστοιχείων. Όσο πιο σωστή είναι η επιλογή των εικονοστοιχείων σήματος και των εικονοστοιχείων υποβάθρου τόσο οι ποσοτικοποιημένες τιμές της έντασης της κάθε κηλίδας ανταποκρίνονται στην πραγματικότητα.

Οι τιμές οι οποίες εξάγονται από την επεξεργασία των εικόνων των μικροσυστοιχιών εν συνεχεία επεξεργάζονται με σκοπό την ανακάλυψη νέας γνώσης. Η πιο συνηθισμένη επεξεργασία είναι η ομαδοποίηση των εντάσεων με σκοπό να ομαδοποιηθούν οι ανιχνευτές που έχουν την ίδια συμπεριφορά στο πείραμα. Σίγουρα τα ποσά της υβριδοποίησης είναι και τα πιο σημαντικά για να καταλήξει κανείς σε συμπεράσματα βιολογικού ενδιαφέροντος. Θα ήταν όμως ενδιαφέρον να μελετηθεί στο μέλλον μήπως και άλλα χαρακτηρίστηκα που μπορούν να εξαχθούν από την επεξεργασία των εικόνων είναι εφικτό να συνεισφέρουν σε χρήσιμα συμπεράσματα. Για παράδειγμα το μέγεθος και το σχήμα της κηλίδας είναι χαρακτηριστικά που προδίδουν τον τρόπο με τον οποίο έλαβε χώρα η υβριδοποίηση. Με τον τρόπο αυτό η επεξεργασία των εικόνων θα τροφοδοτήσει τα εξαγόμενα δεδομένα με παραπάνω στοιχεία τα οποία μπορεί να αποβούν χρήσιμα.

Όσον αφορά τέλος στην αξιοπιστία των μεθόδων για την επεξεργασία των μικροσυστοιχιών, θα ήταν ενδιαφέρουσα η μελέτη της απόδοσης των μεθόδων σε σχέση με τον αριθμό των μη-υβριδοποιημένων ανιχνευτών. Ο αριθμός των ανιχνευτών που δεν

θα υβριδοποιηθούν με τα δείγματα δεν είναι γνωστός και προδιαγεγραμμένος από το πείραμα. Σημαντικό ρόλο διαδραματίζει η στοχευόμενη επιλογή των ανιχνευτών για το εκάστοτε πείραμα. Ωστόσο μια μέθοδος εντοπισμού των κηλίδων στην εικόνα, οφείλει να είναι ακριβείς όσο μεγάλος και να είναι ο αριθμός των μη υβριδοποιημένων κηλίδων στην εικόνα. Είναι καλό λοιπόν να μελετηθεί στο μέλλον πόσες πολλές μπορεί να είναι οι σκοτεινές κηλίδες σε μια εικόνα και πόσο αυτό μπορεί να επηρεάσει την ακρίβεια των μεθόδων.

Περίληψη

Η παρούσα διδακτορική διατριβή στοχεύει στην εξαγωγή βιολογικών δεδομένων από εικόνες που παράγονται από το πείραμα των μικροσυστοιχιών. Σύμφωνα με την υπάρχουσα βιβλιογραφία διάφορες μέθοδοι έχουν προταθεί την επεξεργασία των εικόνων των μικροσυστοιχιών, οι οποίες συνήθως ακολουθούν τα τρία ακόλουθα στάδια:

1. Εντοπισμός των κηλίδων και τοποθέτηση πλέγματος: στοχεύει στον ακριβή εντοπισμό της θέσης κάθε κηλίδας. Η αυτοματοποίηση αυτού του βήματος είναι κρίσιμη για τη διευκόλυνση της ανάλυσης του μεγάλου αριθμού πειραμάτων.
2. Κατάτμηση: διαχωρίζει τα εικονοστοιχεία σήματος από τα εικονοστοιχεία υποβάθρου.
3. Εξαγωγή έντασης: αποτελείται από τον υπολογισμό της μέσης έντασης των κηλίδων σε σχέση με την ένταση του υποβάθρου.

Κατά την διάρκεια της διατριβής υλοποιήθηκε μια αυτόματη μέθοδος για το πρώτο στάδιο της επεξεργασίας, τον εντοπισμό των κηλίδων. Για πρώτη φορά παρουσιάστηκε στην βιβλιογραφία μια μέθοδος, η οποία μπορεί να επεξεργαστεί εικόνες μικροσυστοιχιών με διαφορετική προσέγγιση εκτύπωσης των μικροσυστοιχιών. Δύο είναι οι προσεγγίσεις για την εκτύπωση της μικροσυστοιχίας: (α) ανιχνευτές τοποθετημένοι στις κορυφές τετραγωνικού πλέγματος και (β) ανιχνευτές τοποθετημένοι στις κορυφές εξαγωνικού πλέγματος. Η γενικευμένη μέθοδος η οποία υλοποιήθηκε αρχικά αναγνωρίζει το πλέγμα της εικόνας, και στην συνέχεια υλοποιεί έναν καινοτόμο αλγόριθμο για να εντοπίσει τα κέντρα όλων των ανιχνευτών στην εικόνα. Ο αλγόριθμος αυτός, οποίος ονομάστηκε Αλγόριθμος Ομόκεντρων Αναπτυσσόμενων Πολυγώνων, εκμεταλλεύεται τις κοινές ιδιότητες των δύο πλεγμάτων εντοπίζοντας τα κέντρα επάνω στο περίγραμμα ομόκεντρων τετραγώνων ή εξαγώνων. Επίσης βελτιστοποιεί τις θέσεις των κέντρων που παράγονται. Τέλος η μέθοδος κάνει χρήση του διαγράμματος Βορονόι για να απομονώσει τον κάθε ανιχνευτή.

Στην συνέχεια της διατριβής προχωρήσαμε στην υλοποίηση μιας καινοτόμου μεθόδου για την κατάτμηση (δεύτερο στάδιο επεξεργασίας) των εικόνων με χρήση τεχνικών ταξινόμησης. Η μέθοδος αυτή καινοτομεί σε τρία σημεία, με σκοπό να αποκλείσει από την ποσοτικοποίηση της υβριδοποίησης τα εικονοστοιχεία που προέρχονται από διάφορα τεχνουργήματα της εικόνας. Για τον σκοπό αυτό (α) επεκτάθηκε το διάστημα των χαρακτηριστικών των εικονοστοιχείων, (β) εισήχθη τρίτη κλάση εικονοστοιχείων με σκοπό να λειτουργήσει ως δεξαμενή των εικονοστοιχείων που δεν θέλουμε να συμπεριλάβουμε στις ποσοτικοποιημένες τιμές και τέλος (γ) για πρώτη φορά έγινε τεχνικών ταξινόμησης για τον χαρακτηρισμό των εικονοστοιχείων στις τρεις κλάσεις.

Για την αξιολόγηση των μεθόδων χρησιμοποιήθηκαν εικόνες μικροσυστοιχιών από την βάση δεδομένων μικροσυστοιχιών του Στάνφορντ, εικόνες μικροσυστοιχιών από σταθμό της Illumina, αλλά και προσομοιωμένες εικόνες από των προσομοιωτή του Nykter και των συνεργατών του, ο οποίος χρησιμοποιείται ευρέως στην βιβλιογραφία. Τα αποτελέσματα της διατριβή για τον εντοπισμό των ανιχνευτών παρουσίασαν υψηλά ποσοστά ακρίβεια, τα οποία κυμαίνονται από 92-99%, ενώ η γενικευμένη μέθοδος που υλοποιήθηκε παρουσίασε εξίσου υψηλά ποσοστά με εξειδικευμένες μεθόδους για τις δύο προσεγγίσεις εκτύπωσης. Όσον αφορά στην κατάτμηση επιλέχθηκαν με ακρίβεια τα εικονοστοιχεία σήματος και τα εικονοστοιχεία υποβάθρου, από τα οποία εξήχθησαν οι ποσοτικοποιημένες τιμές. Από την ποσοτικοποίηση εξαιρέθηκαν με επιτυχία εικονοστοιχεία από τεχνουργήματα που εμφανίστηκαν στις εικόνες

Abstract

Microarrays provide a simple way to measure the level of hybridization of known probes of interest with one or more samples under different conditions. Microarray image processing consists of three main stages. The first stage, called spot addressing and gridding, is the procedure for the detection of each spot in the image and the isolation of each spot into a cell. In the second stage, which is known as segmentation stage, each spot of the image is segmented, to separate the signal from the background pixels. Finally, in the third stage the intensity of each spot is extracted and several quantities can be calculated.

For the first stage a generalized method for the spot addressing and the gridding of microarray images is introduced, where the spots are structured either in hexagonal or in rectangular grid. Initially the method identifies the grid of the image. Next, the method utilizes the properties of both the rectangular and the hexagonal grid, to estimate the locations of the non-hybridized spots. This step detects a number of empty spots in each iteration of the proposed Growing Concentric Polygon (GCP) algorithm. The GCP algorithm grows a rectangular or a hexagonal form, for the rectangular structured images or the hexagonal structured images, respectively, detecting a number of spots on the polygon's contour.

For the segmentation stage a novel pixel-by-pixel supervised segmentation method, which is based on classification techniques, is proposed. The method classifies the pixels of the image into three categories using Bayes classifier or Support Vector Machines (SVM). Apart from the signal and the background pixels, the third class includes pixels of artefacts, pixels of the contour of the spot, and pixels of inner holes which exist in donut spots. A set of features from each pixel is used as input for the classification. The proposed method, is advantageous compared to the clustering-based methods, due to the direct characterization of each pixel to the designated category. Otherwise, using clustering techniques different clusters are generated but no distinction exists between them unless a set of rules is applied to separate them.

For the evaluation of the method, both rectangular and hexagonal structured images are employed from the Standard Microarray Database, CNV370 beadchip of Illumina, as well as, simulated images generated by the Nykters simulator. The method results in high accuracy in the detection of the spots ranging from 92 – 99% depending on the dataset used. Additionally, high accuracy results are performed in the segmentation stage, where the signal and background pixels are extracted. Pixels from artefacts are excluded from the intensity extraction stage, to provide more reliable values for the levels of hybridization.

Βιβλιογραφία

- [1] Maton A, Hopkins J, Johnson S, LaHart D, Warner MQ, Wright JD. *Cells Building Blocks of Life*. New Jersey. Prentice Hall, 1997.
- [2] Krout MH. Heredity, environment, and developmental process. *Psychological Review*. 1931;38(3):187-211.
- [3] Keeler CE. Materials for the synthesis of hereditary behavior trends in mammals. *Journal of Comparative and Physiological Psychology*. 1948;41(2):75-81.
- [4] Beres D, Schatia V. Heredity in disease: A review of the recent literature. *The Journal of Pediatrics*. 1938;13(4):555-581.
- [5] Ridley M. *Genome* New York. NY: Harper Perennial. 2006.
- [6] Maton A, Hopkins J, McLaughlin CW, Johnson S, Warner MQ, LaHart D, Wright JD. *Human Biology and Health*. Englewood Cliffs. New Jersey. USA: Prentice Hall. 1993.
- [7] Leuchtenberger C, Leuchtenberger R, Vendrely C, Vendrely R. The quantitative estimation of desoxyribose nucleic acid (DNA) in isolated individual animal nuclei by the caspersson ultraviolet method. *Experimental Cell Research*. 1952;3(1):240-244.
- [8] Goldberg L, Klein E, Klein G. The nucleic acid content of mouse ascites tumor cells. *Experimental Cell Research*. 1950;1(4):543-570.
- [9] Watson JD, Crick FHC. A Structure for Deoxyribose Nucleic Acid. *Nature*. 1953;171:737-738.
- [10] Dumur CI, Dechsukhum C, Ware JL, Cofield SS, Best AM, Wilkinson DS, Garrett CT, Ferreira-Gonzalez A. Genome-wide detection of LOH in prostate cancer using human SNP microarray technology. *Genomics*. 2003;81(3):260-269.
- [11] Jou Y-S, Lo Y-L, Hsiao C-F, Chang G-C, Tsai Y-H, Su W-C, Chen Y-M, Huang M-S, Chen H-L, Chen C-J, Yang P-C, Hsiung CA. Association of an EGFR intron 1 SNP with never-smoking female lung adenocarcinoma patients. *Lung Cancer*. 2009;64(3):251-256.
- [12] Wang Y, Armstrong SA. Genome-Wide SNP Analysis in Cancer: Leukemia Shows the Way. *Cancer Cell*. 2007;11(4):308-309.
- [13] Hemminki K, Bermejo JL. Relationships between familial risks of cancer and the effects of heritable genes and their SNP variants. *Mutation*

- Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2005;592(1-2):6-17.
- [14] Franklin R, Gosling RG. Molecular Configuration in Sodium Thymonucleate. *Nature*.1953;171:740-741.
- [15] Okazaki R, Okazaki T, Sakabe K, Sugimoto K. Mechanism of DNA replication possible discontinuity of DNA chain growth. *Japan Journal in Medicine Science and Biology*. 1967;20(3):255-60.
- [16] Ogawa T, Okazaki T. Discontinuous DNA Replication. *Annu. Rev. Biochem.* 1980; 49:421-457.
- [17] The International HapMap Consortium. Integrating Ethics and Science in the International HapMap Project. *Nature Reviews Genetics*. 2004;5:467-475.
- [18] Thorisson GA, Smith AV, Krishnan L, Stein LD. The International HapMap Project Web site. *Genome Research*. 2005;15:1591-1593.
- [19] Signer ER. Gene expression in foreign cytoplasm. *Journal of Molecular Biology*. 1965;12(1):1-8.
- [20] Matthaei JH, Schoech GK. Human gene expression I: An aminoacyl-RNA binding system from human placenta. *Biochemical and Biophysical Research Communications*. 1967;27(6):638-643.
- [21] Bashiardes S, Lovett M. cDNA detection and analysis. *Current Opinion in Chemical Biology*. 2001;5(1):15-20.
- [22] Harbers M. The current status of cDNA cloning. *Genomics*. 2008;91(3):232-242. 2008.
- [23] Barrett JC, Kawasaki ES. Microarrays: the use of oligonucleotides and cDNA for the analysis of gene expression. *Drug Discovery Today*. 2003;8(3):134-141.
- [24] Yin H, Li S, Zhao X, Du Y, Ma X. cDNA microarray analysis of gene expression in *Brassica napus* treated with oligochitosan elicitor. *Plant Physiology and Biochemistry*. 2006;44(11-12):910-916.
- [25] Gomes LI, Silva RLA, Stolf BS, Cristo EB, Hirata R, Soares FA, Reis LFL, Neves EJ, Carvalho AF. Comparative analysis of amplified and nonamplified RNA for hybridization in cDNA microarray. *Analytical Biochemistry*. 2003;321(2):244-251.

- [26] Touriol C, Bornes S, Bonnal S, Audigier S, Prats H, Prats AC, Vagner S. Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biology of the cell / under the auspices of the European Cell Biology Organization*. 2003;95(3-4):169–78.
- [27] Turanov AA, Lobanov AV, Fomenko DE, Morrison HG, Sogin ML, Klobutcher LA, Hatfield DL, Gladyshev VN. Genetic Code Supports Targeted Insertion of Two Amino Acids by One Codon. *Science*. 2009;323(5911):259-261.
- [28] Viscidi RP, Yolken RG. Molecular diagnosis of infectious diseases by nucleic acid hybridization. *Molecular and Cellular Probes*. 1987;1(1):3-14.
- [29] Cseko YMT, Stone J, Sederoff R. Nucleic acid hybridization of highly repeated DNA in extracts of single *Drosophila*. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*. 1979;565(2):253-264.
- [30] Bartlett J, Stirling D. A Short History of the Polymerase Chain Reaction. *Methods Mol Biol*. 2003;226:3-6.
- [31] Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*. 1985;230(4732):1350-1354.
- [32] Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*. 1988;239:487-491.
- [33] Cheng S, Fockler C, Barnes WM, Higuchi R. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proceedings of Natl Acad Sci*. 1994;91:5695-5699.
- [34] Sambrook J, Russel DW. *Molecular Cloning: A Laboratory Manual* (3rd ed.). Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press, 2001.
- [35] Pavlov AR, Pavlova NV, Kozyavkin SA, Slesarev AI. Recent developments in the optimization of thermostable DNA polymerases for efficient applications. *Trends Biotechnology*. 2004;22:253-260.
- [36] Rychlik W, Spencer WJ, Rhoads RE. Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Research*. 1990;18:6409-6412.

- [37] Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. *Science*. 1991;251(4995):767-773.
- [38] Schena M, Shalon D, Davis RW, Brown PO. Quantitative motoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995;270:467-470.
- [39] Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol*. 1996;14(13):1675-80.
- [40] Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proceedings of the National Academy of Sciences*. 1997;94(24):13057-13062.
- [41] Shen R, Fan JB, Campbell D, Chang W, Chen J, Doucet D, Yeakley J, Bibikova M, Wickham GE, McBride C, Steemers F, Garcia F, Kermani BG, Gunderson K, Oliphant A. High-throughput SNP genotyping on universal bead arrays. *Mutation Research*. 2005;573:70-82.
- [42] Fernandez-Teijeiro A, Betensky RA, Sturla LM, Kim JY, Tamayo P, Pomeroy SL. Combining gene expression profiles and clinical parameters for risk stratification in medulloblastomas. *Journal of Clinical Oncology*. 2004;22(6):994-998.
- [43] Hoffman EP, Awad T, Palma J, Webster T, Hubbell E, Warrington JA, Spira A, Wright G, Buckley J, Triche T, Davis R, Tibshirani R, Jones WXW, Tompkins R, West M. Expression profiling - best practices for data generation and interpretation in clinical trials. *Nature Reviews Genetics*. 2004;5:229-237.
- [44] Rhodes DR, Chinnaiyan AM. Bioinformatics strategies for translating genome-wide expression analyses into clinically useful cancer markers. *Annals of New York Academy of Science*. 2004;1020:32-40.
- [45] Fang Y, Lahiri J, Picard L. G protein-coupled receptor microarrays for drug discovery. *Drug Discovery Today*. 2003;8(16):755-761.

- [46] Clarke PA, Poele R, Wooster R, Workman P. Gene expression microarray analysis in cancer biology, pharmacology, and drug development: progress and potential. *Biochemical Pharmacology*. 2001;62(10):1311-1336.
- [47] Bailey SN, Wu RZ, Sabatini DM. Applications of transfected cell microarrays in high-throughput drug discovery. *Drug Discovery Today*. 2002;7(18):S113-S118.
- [48] Kim H-J, Park S-H, Lee T-H, Nahm B-H, Kim Y-R, Kim H-Y. Microarray detection of food-borne pathogens using specific probes prepared by comparative genomics. *Biosensors and Bioelectronics*. 2002;24(2):238-246.
- [49] Liu-Stratton Y, Roy S, Sen CK. DNA microarray technology in nutraceutical and food safety. *Toxicology Letters*. 2004;150(1):29-42.
- [50] Schmidt-Heydt M, Geisen R. A microarray for monitoring the production of mycotoxins in food. *International Journal of Food Microbiology*. 2004;117(2):131-140.
- [51] Shinawi M, Cheung SW. The array CGH and its clinical applications. *Drug Discov Today*. 2008;17-18:760-70.
- [52] MacBeath G, Schreiber SL. Printing Proteins as Microarrays for High-Throughput Function Determination. *Science*. 2000;289(5485):1760-1763.
- [53] Morozov VN. Protein microarrays: Principles and limitations. *Protein Microarrays*. Schena M. Ed. Jones and Bartlett Publishers. Inc. Sudbury MA; 2005, p. 71–106.
- [54] Wang XH, Istepanian RSH, Song YH. Microarray image enhancement by denoising using stationary wavelet transform. *IEEE Transactions on Nanobioscience*. 2003;2:184-189.
- [55] Shalon D, Smith SJ, Brown PO. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Research*. 1996;6(7):639-645.
- [56] Weibel C. The Spotting Accelerator™, customizable head assembly for advanced microarraying. *Journal of the Association for Laboratory Automation*. 2002;7(3):91-96.
- [57] University of Stanford. 2005. <http://cmgm.stanford.edu/pbrown/mguide/tips.html>.
- [58] Xia YN, Whitesides GM. Soft lithography. *Angewandte Chemie-International Edition*. 1998;37(5):551-575.

- [59] Zhao XM, Xia YN, Whitesides GM. Soft lithographic methods for nanofabrication. *Journal of Materials Chemistry*. 1997;7(7):1069-1074.
- [60] Lee KB, Park SJ, Mirkin CA, Smith JC, Mirkovich M. Protein nanoarrays generated by dip-pen nanolithography. *Science*. 2002;295(5560):1702-1705.
- [61] Piner RD, Zhu J, Xu F, Hong SH, Mirkin CA. "Dip-pen" nanolithography. *Science*. 1999;283(5402):661-663.
- [62] Blawas AS, Reichert WM. Protein patterning. *Biomaterials*. 1998;19(7-9):595-609.
- [63] Lom B, Healy KE, Hockberger PE. A versatile technique for patterning biomolecules onto glass coverslips. *Journal of Neuroscience Methods*. 1993;50(3):385-397.
- [64] Mooney JF, Hunt AJ, McIntosh JR, Liberko CA, Walba DM, Rogers CT. Patterning of functional antibodies and other proteins by photolithography of silane monolayers. *Proceedings of the National Academy of Sciences of the United States of America*. 1996;93(22):12287-12291.
- [65] Pirrung MC, Huang CY. A general method for the spatially defined immobilization of biomolecules on glass surfaces using "caged" biotin. *Bioconjugate Chemistry*. 1996;7(3):317-321.
- [66] Hengsakul M, Cass AEG. Protein patterning with a photoactivatable derivative of biotin. *Bioconjugate Chemistry*. 1996;7(2):249-254.
- [67] Pritchard DJ, Morgan H, Cooper JM. Patterning and regeneration of surfaces with antibodies. *Analytical Chemistry*. 1995;67(19):3605-3607.
- [68] Zeng J, Deshpande M, Kan H-C, Gilbert JR. A dynamic spotting method for split-pin based microarrays. *Micro Total Analysis Systems (MicroTAS) Conference*. 2001;143-144.
- [69] Allain LR, Askari M, Stokes DL, Vo-Dinh T. Microarray sampling-platform fabrication using bubble-jet technology for a biochip system. *Fresenius Journal of Analytical Chemistry*. 2001;371(2):146-150.
- [70] Tseng F-G, Lin S-C, Yao D-J, Huang H, Chieng C-C. Technological aspects of protein microarrays and nanoarrays. *Protein Microarrays*. Schena M. Ed. Jones and Bartlett Publishers Inc.; 2005. p. 305-338.

- [71] Chrisey DB, McGill RA, Pique A. Matrix assisted pulsed laser evaporation direct write. U. S. Patent. 2001; 6:177-181,
- [72] Ringeisen BR, Wu PK, Kim H, Pique A, Auyeung RYC, Young HD, Chrisey DB, Krizman DB. Picoliter-scale protein microarrays by laser direct write. *Biotechnology Progress* 2002;18(5):1126-1129.
- [73] Ernst LA, Gupta RK, Mujumdar RB, Waggoner AS. Cyanine dye labeling reagents for sulfhydryl groups. *Cytometry*. 1989;10(1):3-10.
- [74] Gollub J, Ball CA, Binkley G, Demeter, Finkelstein DB, Hebert JM, Hernandez-Boussard T, Jin H, Kaplper M, Matese JC, Schroeder M, Brown PO, Botstein D, Sherlock G. The Stanford Microarray Database: data access and quality assessment tools. *Nucleic Acids Research*. 2003;31:94-96.
- [75] Eisen MB. ScanAlyse, <http://rana.Stanford.EDU/software/>. 1999.
- [76] Fielden MR, Halgren RG, Dere E, Zacharewski TR. GP3: GenePix post-processing program for automated analysis of raw microarray data. *Bioinformatics*. 2002;18(5):771-773.
- [77] Illumina inc.: <http://www.illumina.com/>.
- [78] Nykter M, Aho T, Ahdesmäki M, Ruusuvoori P, Lehmissola A, Yli-Harja O. Simulation of microarray data with realistic characteristics. *BMC Bioinformatics*. 2006;7(1):349.
- [79] Schena M. *Microarray Biochip Technology*. Eaton Publishing Company/Biotechniques Books. 1 edition. 2000.
- [80] Bajcsy P. An overview of DNA Microarray Image Requirements for Automated Processing. *IEEE Conference on Computer Vision and Pattern Recognition*. 2005;3:147.
- [81] Dudoit S, Yang YH, Callow MJ, Speed TP. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica*. 2002;12:111-139.
- [82] Eisen MB, Brown PO. DNA arrays for analysis of gene expression. *Methods of Enzymology*. 1999;303:179-205.
- [83] Yang YH, Buckley MJ, Speed TP. Analysis of cDNA microarray images. *Briefings in bioinformatics*. 2001;2(4):341-349.

- [84] Hautaniemi S, Lehmussola A, Yli-Harja O. DNA microarray data preprocessing. First International Symposium on Control. Communications and Signal Processing. 2004;751-754.
- [85] Brandle N, Chen H-Y, Bischof H, Lapp H. Robust parametric and semi-parametric spot fitting for spot array images. 8th International Conference on Intelligent Systems for Molecular Biology. 2000;46-56.
- [86] Glasbey CA, Ghazal P. Combinatorial image analysis of DNA microarray features. *Bioinformatics*. 2003;19(2):194-203.
- [87] Wang L-Q, Ni X-X, Lu Z-K, Zheng X-F, Li Y-S. Enhancing the quality metric of protein microarray image. *J. Zhejiang Univ. Science*. 2004;5(12):1621-1628.
- [88] Bengtsson A, Bengtsson H. Microarray image analysis: background estimation using quantile and morphological filters. *BMC Bioinformatics*. 2006;7:96.
- [89] Bozinov D, Seidel P-M. Iterative gridding for automated microarray image analysis. 38th Asilomar conference on signals systems and computers. 2004;2:1635-1640.
- [90] Lukac R, Plataniotis KN, Smolka B, Venetsanopoulos AN. A multichannel order-statistic technique for cDNA microarray image processing. *IEEE Transactions on NanoBioscience*. 2004;3(4):272-285.
- [91] Lukac R, Plataniotis KN. Fuzzy Vector Filters for cDNA Microarray Image Processing. *Fuzzy Systems in Bioinformatics and Computational Biology*. 2009;242:67-82.
- [92] Steinfath M, Wruck W, Seidel H, Lehrach H, Radelof U, O'Brien J. Automated image analysis for array hybridization experiments. *Bioinformatics*. 2001;17:634-641.
- [93] Deng N, Duan H. An Automatic and Power Spectra-based Rotate Correcting Algorithm for Microarray Image. in Proc of 27th Annual International Conference of the Engineering in Medicine and Biology Society. 2005;898-901.
- [94] Bowman C, Baumgartner R, Booth S. Automated Analysis of Gene-microarray Images. in Proc. IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). 2002;1140-1144.

- [95] Deepa J, Tessamma T. Automatic Gridding of DNA Microarray Images using Optimum Subimage. *International Journal of Recent Trends in Engineering*. 2009;1(4):37-40.
- [96] Fabbri R, Costa LF, Barrera J. Towards Non-Parametric Gridding of Microarray Images. *14th International Conference on Digital Signal Processing*. 2002;2(1-3):623-626.
- [97] Jain AN, Tokuyasu TA, Snijders AM, Segraves R, Albertson DG, Pinkel D. Fully Automated Quantification of Microarray Image Data. *Genome Research*. 2002;12(2):325-332.
- [98] Bajcsy P. Gridline: Automatic Grid Alignment in DNA Microarray Scans. *IEEE Trans. on Image Processing*. 2004;13(1):15–25.
- [99] Angulo J, Serra J. A mathematical morphology contribution to the analysis of DNA microarray images. *5th ESMTB Conference Mathematical Modelling & Computing in Biology and Medicine*. Milano. Italy. 2002.
- [100] Angulo J, Serra J. Automatic analysis of DNA microarray images using mathematical morphology. *Bioinformatics*. 2003;19:553-562.
- [101] Hirata R, Barrera J, Hashimoto RF, Dantas D. Microarray gridding by mathematical morphology. in *Proceedings of the Ijth Brazilian Symposium on Computer Graphics and Image Processing*. Brazil. 2001;112-119.
- [102] Hirata R, Barrera J, Hashimoto RF, Dantas DO, Esteves GH. Segmentation of microarray images by mathematical morphology. *Real-Time Imaging*. 2002;8:491-505.
- [103] Baek J, Son YS, McLachlan GJ. Segmentation and intensity estimation of microarray images using a gamma-t mixture model. *Bioinformatics*. 2007;23(4):458-465.
- [104] Lonardi S, Luo Y. Gridding and compression of microarray images. in *Proc. of IEEE Computational Systems Bioinformatics Conference*. Stanford. CA. 2004;122-130.
- [105] Bariamis D, Iakovidis DK, Maroulis D. M3G: Maximum Margin Microarray Gridding. *BMC Bioinformatics*. 2010;11:49.

- [106] Bariamis D, Maroulis DE, Iakovidis DK. Unsupervised SVM-based Gridding for DNA Microarray Images. *Computerized Medical Imaging and Graphics*, Elsevier, article in press.
- [107] Bariamis D, Maroulis D, Iakovidis DK. Automatic DNA Microarray Gridding based on Support Vector Machines. in *Proc. IEEE International Conference on Bioinformatics and Biotechnology*. 2008.
- [108] Zacharia E, Maroulis D. An original Genetic Approach to the Fully-Automatic Gridding of Microarray Images. *IEEE Trans. on Medical Imaging*. 2008;27(6):805-813.
- [109] Zacharia E, Maroulis D. Microarray image gridding via an evolutionary algorithm. in *Proc of IEEE International Conference on Image Processing (ICIP2008)*. 2008;1444-1447.
- [110] Jung H-Y, Cho H-G. An automatic block and spot indexing with k-nearest neighbors graph for microarray image analysis. *Bioinformatics*. 2002;18(Suppl. 2):S141–S151.
- [111] Chen W-B, Zhang C, Liu W-L. An Automated Gridding and Segmentation Method for cDNA Microarray Image Analysis. *19th IEEE International Symposium on Computer-Based Medical Systems*. 2006;893-898.
- [112] Otsu N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*. 1979;9:62-66.
- [113] Brunelli R. *Template Matching Techniques in Computer Vision: Theory and Practice*. Wiley. 2009.
- [114] Galinsky VL. Automatic registration of microarray images. I. Rectangular grid. *Bioinformatics*. 2003;19:1824-1831.
- [115] Galinsky VL. Automatic registration of microarray images. II. Hexagonal grid. *Bioinformatics*. 2003;19:1832-1836.
- [116] Ceccarelli M, Antoniol G. A Deformable Grid-Matching Approach for Microarray Images. *IEEE Trans. on Image Proces*. 2006;15(10):3178-3188.
- [117] Bidaut G, Manion F, Garcia C, Ochs M. WaveRead: Automatic Measurement of Relative Gene Expression Levels from Microarrays using Wavelet Analysis. *Journal of Biomedical Informatics*. 2006;39(4):379-388.

- [118] Hartelius K, Cartstensen JM. Bayesian grid matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2003;25(2):162-173.
- [119] Preparata FP, Shamos MI. *Computational geometry: an introduction*. Texts and Monographs in Computer Science. Springer. Berlin. 1985.
- [120] Aurenhammer F. Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure. in *ACM Comp. Surveys*. 1991;23:345-405.
- [121] Antoniol G, Ceccarelli M. Microarray image addressing based on the radon transform. *Proceedings of IEEE International Conference on Image Processing*. Washington. USA. 2005;13-16.
- [122] Ho J, Hwang WL, Lu HHS, Lee DT. Gridding Spot Centers of Smoothly Distorted Microarray Images. *IEEE Transaction on Image Processing*. 2006;15(2):342-353.
- [123] Zwillinger D. *Affine Transformations*. CRC Standard Mathematical Tables and Formulae. Boca Raton. FL: CRC Press; 1995. p. 265-266.
- [124] Shapiro LG, Stockman GC. *Computer Vision*. Englewood Cliffs. NJ: Prentice-Hall. 2001.
- [125] Deans SR. *The Radon Transform and Some of Its Applications*. New York: Wiley. 1983.
- [126] Ahmed AA, Vias M, Iyer NG, Caldas C, Brenton JD. Microarray segmentation methods significantly influence data precision. *Nucleic Acids Research*. 2004;32:e50.
- [127] Cathelin R, Lopez F, Klopp C. AGScan: a pluggable microarray image quantification software based on the Image J library. *Bioinformatics*. 2007;23:247-248. 2007.
- [128] Liew AW-C, Yang H, Yang M. Robust Adaptive Spot Segmentation of DNA Microarray Images. *Pattern Recognition*. 2003;36:1251-1254.
- [129] Buhler J, Ideker T, Haynor D. Dapple: improved techniques for finding spots on DNA microarrays. *UWCSE Tech Rep. UWTR Dept. of Comp. Science and Eng. University of Washington*. 2000.
- [130] Axon Instruments. *GenePix A User's Guide*. 1999.

- [131] Rueda LG, Qin L. A New Method for DNA Microarray Image Segmentation in proc of International Conference on Image Analysis and Recognition. 2005;886-893.
- [132] Chen Y, Dougherty ER, Bittner ML. Ratio-Based Decisions and the Quantitative Analysis of cDNA Microarray Images. *Journal of Biomedical Optics*. 1997;2(4):364-374.
- [133] Hua J, Xiong Z, Wu Q, Castleman K. Fast segmentation and lossy-to-lossless compression of DNA microarray images. in Proc. Workshop on Genomic Signal Processing and Statistics. Raleigh. NC. USA. 2002.
- [134] Hua J, Liu Z, Xiong Z, Wu Q, Castleman KR. Microarray BASICA: Background adjustment, segmentation, image compression and analysis of microarray images. In *EURASIP Journal on Applied Signal Processing: Special Issue on Genomic Signal Processing*. 2004;92-107.
- [135] GSI Lumonics (1999). QuantArray Analysis Software. Operator's Manual.
- [136] QuantArray Analysis Software. <http://lifesciences.perkinelmer.com>.
- [137] Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*. 1947;18:50-60.
- [138] Adams R, Bischof L. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1994;16:641-647.
- [139] Jouenne VY. Critical Issues in the Processing of cDNA Microarray Images. Master of Science. Virginia Polytechnic Institute and State University. 2001.
- [140] Yang YH, Buckley MJ, Dudoit S, Speed TP. Comparison of Methods for Image Analysis on cDNA Microarray Data. Technical Report #584. Department of Statistics. University of California at Berkeley. 2000.
- [141] Lehmußola A, Ruusuvoori P, Yli-Harja O. Evaluating the performance of microarray segmentation algorithms. *Bioinformatics*. 2006;22(23):2910-2917.
- [142] Roerdink JBTM, Meijster A. The watershed transform: definitions, algorithms and parallelization strategies. *Fundamenta Informaticae*. 2000;41(1-2):187-228.
- [143] Park C-B, Lee K-W, Lee S-W. Automatic Microarray Image Segmentation Based on Watershed Transformation. 17th International Conference on Pattern Recognition (ICPR'04). 2004;3:786-789.

- [144] Shih C-L, Chiu H-W. Automatic spot detection of cDNA Microarray images using mathematical morphology methods. Conference on IEEE EMBS Asian-Pacific. 2003;70-71.
- [145] Siddiqui KI, Hero AO, Siddiqui MM. Mathematical morphology applied to spot segmentation and quantification of gene microarray images. in Conference on Conference Record of the Thirty-Sixth Asilomar. 2002;1:926-930.
- [146] Kornaros G. A soft multi-core architecture for edge detection and data analysis of microarray images. Journal of Systems Architecture. 2010;56(1):48-62.
- [147] Sobel I, Feldman G. An isotropic 3x3 image gradient operator. in H. Freeman (Ed.) Machine Vision for Three-dimensional Scenes. Academic Press. New York. 1990;376-379.
- [148] Kim JH, Kim HY, Lee YS. A novel method using edge detection for signal extraction form cDNA microarray image analysis. Experimental and Molecular Medicine. 2001;33:83-88.
- [149] Srinark T, Kambhamettu C. A microarray image analysis system based on multiple snakes. J. Biol. Syst. Special Issue. 2004;12:127- 157.
- [150] Gjerstad Ø, Aakra Å, Indahl U. Modelling and quality assessment of spots in digital Images - with application to DNA microarrays. Chemometrics and Intelligent Laboratory Systems. 2009;98:10-23.
- [151] Gjerstad Ø, Aakra Å, Snipenand L, Indahl U. Probabilistically assisted spot segmentation - With application to DNA microarray images. Chemometrics and Intelligent Laboratory Systems. 2009;98(1):1-9.
- [152] MacQueen JB. Some Methods for classification and Analysis of Multivariate Observations. in 5-th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley. University of California Press. 1967;1:281-297.
- [153] Battiato S, Blasi GDi, Farinella GM, Gallo G, Guarnera GC. Ad-Hoc Segmentation Pipeline for Microarray Image Analysis. in IS&T-SPIE Electronic Imaging. San Jose. USA. 2006.
- [154] Bozinov D, Rahnenführer J. Unsupervised technique for robust target separation and analysis of DNA microarray spots through adaptive pixel clustering. Bioinformatics. 2002;18:747-756.

- [155] Ergüt E, Yardimci Y, Mumcuoglu E, Konu O. Analysis of microarray images using FCM and K-means clustering algorithm. In IJCI 2003. 2003;116-121.
- [156] Wu H, Yan H. Microarray Image Processing Based on Clustering and Morphological Analysis. Proc. of First Asia Pacific Bioinformatics Conference. Adelaide. Australia. 2003;111-118.
- [157] Nagarajan R. Intensity-Based Segmentation of Microarray Images. IEEE Trans. On Medical Imaging. 2003;22(7):882-889.
- [158] Nagarajan R, Peterson CA. Identifying Spots in Microarray Images. IEEE Transactions on Nanobioscience. 2002;1(2):78-84.
- [159] Wang Y-P, Gunampally MR, Cai W-W. Automated segmentation of microarray spots using fuzzy clustering approaches. IEEE International Workshop on Machine learning for signal processing. 2005;387-391.
- [160] Uslan V, Bucak IÖ. Microarray image segmentation using clustering methods. Mathematical and Computational Applications. 2010;15(2):240-247.
- [161] Bezdek JC. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press. New York. 1981.
- [162] Demirkaya O, Asyali MH, Shoukri MM, Abu-Khabar KS. Segmentation of Microarray cDNA Spots Using MRF-Based Method. Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2003;674-677.
- [163] Demirkaya O, Asyali MH, Shoukri MM. Segmentation of cDNA Microarray Spots Using Markov Random Field Modeling. Bioinformatics. 2005;21(13):2994-3000.
- [164] Katzer M, Kummert F, Sagerer G. Methods for automatic microarray image segmentation. IEEE Transactions on Nanobioscience. 2003;2:202-214.
- [165] Katzer M, Kummert F, Sagerer G. A Markov Random Field Model of Microarray Gridding. in Proc. ACM Symposium on Applied Computing (SAC). (Melbourne, Florida). 2003;72-77.
- [166] Kaufman L, Rousseeuw PJ. Finding Groups in Data - An Introduction to Cluster Analysis. Wiley. NY. 1989.
- [167] Athanasiadis EI, Cavouras DA, Spyridonos PP, Glotsos DTh, Kalatzis IK, Nikiforidis G. Complementary CNA Microarray Image Processing Based on the

- Fuzzy Gaussian Mixture Model. *IEEE Transactions on Information Technology in Biomedicine*. 2009;13(4):419-425.
- [168] Blekas K, Galatsanos NP, Georgiou I. An Unsupervised Artifact Correction Approach for the Analysis of DNA Microarray Images. in *Proc. IEEE International Conf. on Image Processing (ICIP)*. 2003;2:165-168.
- [169] Blekas K, Galatsanos N, Likas A, Lagaris IE. Mixture Model Analysis of DNA Microarray Images. *IEEE Transactions on Medical Imaging*. 2005;24(7):901-909.
- [170] Tom BDM, Gilks WR, Brooke-Powell ET, Ajioka JW. Quality determination and the repair of poor quality spots in array experiments. *BMC Bioinformatics*. 2005;6:234.
- [171] Chen T-B, Horng-Shing H, Lee L-S, Lan H-J. Segmentation of cDNA microarray images by kernel density estimation. *Journal of Biomedical Informatics*. 2008;41(6): 1021-1027.
- [172] Liew AW-C, Yan H. Automated cDNA microarray image segmentation. *Proceedings of the International Symposium on Computational Models for Life Sciences*. 2007.
- [173] Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1977;39(1):1-38.
- [174] Bozinov D. Autonomous system for web-based microarray image analysis. *IEEE Trans. Nanobiosci*. 2003;2:215-220.
- [175] Rahnenfuhrer J, Bozinov D. Hybrid clustering for microarray image analysis combining intensity and shape features. *BMC Bioinformatics*. 2004;5:47.
- [176] Abbaspour M, Abugharbieh R, Podder M, Tripp B, Tebbutt S. Hybrid Spot Segmentation in Four-Channel Microarray Genotyping Image Data. *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. Vancouver, Canada. 2006;11-16.
- [177] Sonca M, Hlavac V, Boyle R. *Image Processing, Analysis, and Machine Vision*. (second edition). PWS Publishing. United States of America. 1999. ISBN: 0-534-95393-X.
- [178] Grubbs FE. Procedures for detecting outlying observations in samples. *Technometrics*. 1969;11:1-21.

- [179] Giannakeas N, Fotiadis DI, Politou AS. An Automated Method for Gridding in Microarray Images. 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2006;5876-5879. New York. USA.
- [180] Giannakeas N, Fotiadis DI. An Automated Method for Gridding and Segmentation Of cDNA Microarray Images. Computerized Medical Imaging and Graphics. 2009;33(1):40-49.
- [181] Kalatzis FG, Giannakeas N, Exarchos TP, Lorenzelli L, Adami A, Decarli M, Lupoli S, Macciardi F, Markoula S, Georgiou I, Fotiadis DI. Developing a genomic-based point-of-care diagnostic system for rheumatoid arthritis and multiple sclerosis. 31th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2009:827-830. Mineapolis. U.S.A.
- [182] Giannakeas N, Kalatzis T, Fotiadis DI. Spot Addressing for Microarray Images Structured in Hexagonal Grids. Computer Methods and Programs in Biomedicine. 2010. in press.
- [183] Giannakeas N, Kalatzis T, Fotiadis DI. A Generalized Method for the Gridding of Microarray Images with rectangular or hexagonal grid. 2011, in Press
- [184] Ionita-Laza I, Rogers AJ, Lange C, Raby BA, Lee C. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. Genomics. 2009;93(1).
- [185] Smyth GK, Yang YH, Speed T. Statistical Issues in cDNA Microarray Data Analysis. Functional Genomics: Methods and Protocols. MJ Brownstein and AB Khodursky (ed.) Methods in Molecular Biology series Humana Press Totowa NJ. 2002.
- [186] Giannakeas N, Fotiadis DI. Image Processing and Machine Learning Techniques for the Segmentation of cDNA Microarray Images. In T Exarchos, A Papadopoulos, DI Fotiadis (Eds.). Handbook of research on advanced techniques in diagnostic images and biomedical application. 2009;294-306. New York:IGI.
- [187] Giannakeas N, Fotiadis DI .Multichannel Segmentation of cDNA Microarray Images using the Bayes Classifier. 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2007;3466-3469. Lyon. France.

- [188] Giannakeas N, Karvelis PS, Fotiadis DI. A Classification-Based Segmentation of cDNA Microarray Images using Support Vector Machines. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2008;875-878 Vancouver. Canada.
- [189] Giannakeas N, Karvelis PS, Exarchos TP, Kalatzis FG, Fotiadis DI. Segmentation of Microarray Images Using Pixel Classification - Comparison with Clustering based Methods. Artificial Intelligent in Medicine. 2010. in Press.
- [190] Jolliffe IT. Principal Component Analysis. Series: Springer Series in Statistics, 2nd ed. Springer. NY. XXIX. 2002:487.
- [191] Li Q, Fraley C, Bumgarner RE, Yeung KY, Raftery AE. Donuts, scratches and blanks: robustmodel-based segmentation of microarray images. Bioinformatics. 2005;21:2875-82.
- [192] Gonzalez RC, Woods RE, Eddins SL. Digital image processing using MATLAB. Prentice Hall. Upper Saddle River. NJ. 2004.
- [193] Burges CJC. A Tutorial on Support Vector Machines for Pattern Recognition. D. M. and Kn. Discovery. 1998;2:121-167.