



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ
ΥΛΙΚΩΝ

**ΑΥΤΟΜΑΤΗ ΕΦΑΡΜΟΓΗ ΠΛΕΓΜΑΤΟΣ ΣΕ ΕΙΚΟΝΕΣ DNA
(DNA MICROARRAYS)
ΜΕ ΧΡΗΣΗ ΤΕΧΝΙΚΩΝ ΨΗΦΙΑΚΗΣ ΕΠΕΞΕΡΓΑΣΙΑΣ
ΕΙΚΟΝΩΝ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Χριστόφορος Χ. Χαραλάμπους

Επιβλέπων: Γιώργος Ματσόπουλος
Επ. Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2011



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ
ΥΛΙΚΩΝ

**ΑΥΤΟΜΑΤΗ ΕΦΑΡΜΟΓΗ ΠΛΕΓΜΑΤΟΣ ΣΕ ΕΙΚΟΝΕΣ DNA
(DNA MICROARRAYS)
ΜΕ ΧΡΗΣΗ ΤΕΧΝΙΚΩΝ ΨΗΦΙΑΚΗΣ ΕΠΕΞΕΡΓΑΣΙΑΣ
ΕΙΚΟΝΩΝ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Χριστόφορος Χ. Χαραλάμπους

Επιβλέπων: Γιώργος Ματσόπουλος
Επ. Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 5^η Ιουλίου 2011.

.....
Γ. Ματσόπουλος
Επ. Καθηγητής Ε.Μ.Π.

.....
Ν. Ουζούνογλου
Καθηγητής Ε.Μ.Π.

.....
Κ. Νικήτα
Καθηγήτρια Ε.Μ.Π.

Αθήνα, Ιούλιος 2011

.....
Χριστόφορος Χ. Χαραλάμπους

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Χριστόφορος Χ. Χαραλάμπους, 2011

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Σκοπός της παρούσας διπλωματικής εργασίας είναι η παρουσίαση της μεθόδου αυτόματης εφαρμογής πλέγματος σε εικόνες DNA και σύγκριση της με κάποιες από τις υπάρχουσες μεθόδους που χρησιμοποιούνται.

Προκειμένου να εξαχθούν κάποια ασφαλή συμπεράσματα, οι εικόνες αυτές θα πρέπει να υποστούν μια σειρά από βήματα επεξεργασίας ώστε να μπορούν εύκολα να αναλυθούν περαιτέρω. Ο τρόπος εξαγωγής και ψηφιοποίησης τους αλλά και η σειρά των βιολογικών διαδικασιών που έλαβαν χώρα από την λήψη των δειγμάτων μέχρι και την εκτύπωση των εικόνων βοηθούν στην αναγνώριση των παραγόντων που είναι σημαντικοί για την ανάλυση τους. Σημαντικό ρόλο παίζει η ύπαρξη ή όχι των στιγμάτων καθώς επίσης και η φωτεινότητα τους.

Τεχνικές κατάτμησης είναι απαραίτητες, ώστε το κάθε ένα στίγμα να μπορεί να υποστεί την σωστή επεξεργασία για ανάλυση του περιγράμματος και της φωτεινότητας του. Για να είναι μια τέτοια τεχνική κατάτμησης εφικτή, θα πρέπει πρώτα να εφαρμοστεί ένα πλέγμα (grid) πάνω στην εικόνα ώστε να γίνεται εύκολος ο διαχωρισμός των στιγμάτων και η άμεση αναφορά σε συγκεκριμένα στίγματα γνωστών συντεταγμένων.

Με βάση αυτά, αναδεικνύεται η ανάγκη μιας μεθόδου αυτόματης εφαρμογής ενός πλέγματος που να ταιριάζει με τις εικόνες οι οποίες μπορεί να έχουν υποστεί παραμορφώσεις διαφόρων μορφών. Στα πλαίσια της διπλωματικής εργασίας, υλοποιήθηκε μια αυτόματη μέθοδος, η οποία μπορεί να αντεπεξέλθει επιτυχώς σε τέτοιες δυσμενείς συνθήκες και να δώσει πολύ ικανοποιητικά αποτελέσματα.

Λέξεις κλειδιά

Εικόνες DNA, πλέγμα, ευθυγράμμιση, ταίριασμα προτύπων, συντελεστές συσχέτισης, ανίχνευση στιγμάτων, επεξεργασία εικόνας.

Abstract

The objective of the specific thesis is the presentation of a methodology for automatic gridding of DNA microarrays and the qualitative and quantitative comparison of it, with some of the methodologies that are used today.

In order to draw any reliable conclusions, these DNA images may have to be subjected to a series of preprocessing steps so they can easily further analyzed. The way these experiments had taken place and been digitalized, as well as the range of biological processes that took place following the reception of the samples, until the printing of the images, help in identifying the factors that are important to analyze these images. The existence or not of the spots, as well as the hybridization of them, play a significant role to the whole data analysis.

Segmentation techniques are necessary, so that each spot can be subjected to proper processing treatment for analysis of its contour and brightness. In order for such a technique to be feasible, a grid should first be applied on the image so to separate the spots from each other and have direct reference to specific spots of known coordinates.

According to all the above, the need for an automatic method for gridding DNA microarrays, which may have undergone various forms of distortion is highlighted. As part of this thesis, an automatic method for gridding DNA microarrays was implemented. This method can successfully withstand such adverse conditions and bring very satisfying results.

Key words

DNA microarrays, grid, alignment, template matching, correlation coefficients, spots detection, image processing.

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε κατά το ακαδημαϊκό έτος 2010-2011 υπό την επίβλεψη του κ. Γεώργιου Ματσόπουλου, επ. καθηγητή του Ε.Μ.Π. της σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, στον οποίο οφείλω ιδιαίτερες ευχαριστίες για την ανάθεσή της, δίνοντας μου την ευκαιρία να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα. Επίσης, θα ήθελα να ευχαριστήσω τον Δρ. Νικόλαο Μουραβλιάνσκι για τη βοήθεια και την καθοδήγηση που μου παρείχε κατά την εκπόνηση της εργασίας. Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου, την αδερφή μου και τους φίλους μου για τη στήριξη και βοήθεια τους καθ' όλη τη διάρκεια των σπουδών μου.

Περιεχόμενα

Εισαγωγή	- 1 -
1.1. Σκοπός Διπλωματικής Εργασίας	- 1 -
1.2. Διάρθρωση Διπλωματικής Εργασίας	- 2 -
Εικόνες DNA (DNA microarrays)	- 5 -
2.1. Εισαγωγή στα γονίδια, DNA και πρωτεΐνες	- 5 -
2.2. Περιγραφή της τεχνολογίας των εικόνων DNA (DNA microarrays technology).....	- 7 -
2.3. Διαδικασία δημιουργίας των εικόνων DNA	- 8 -
2.3.1. Συλλογή Δεδομένων	- 8 -
2.3.2. Απομόνωση του αγγελιαφόρου RNA (mRNA)	- 8 -
2.3.3. Επισήμανση του mRNA	- 8 -
2.3.4. Υβριδοποίηση	- 8 -
2.3.5. Καθαρισμός Εικόνων και Μέτρηση Φωτεινότητας	- 9 -
2.4. Διαδικασία Ανάγνωσης Εικόνων DNA	- 10 -
2.4.1. Ανάγνωση Ανεπεξέργαστων Δεδομένων	- 10 -
2.4.2. Εφαρμογή πλέγματος (gridding process).....	- 11 -
2.4.3. Αξιολόγηση της Ποιότητας των Δεδομένων	- 11 -
2.4.4. Κβαντοποίηση – Κανονικοποίηση	- 11 -
2.4.5. Προσαρμογή των Αποτελεσμάτων σε Διεθνή Πρότυπα.....	- 12 -
Τεχνικές Εφαρμογής Πλέγματος.....	- 13 -
3.1. Διαδικασία Εφαρμογής Πλέγματος – Αρχική Προσέγγιση	- 13 -
3.2. Υπάρχουσες Τεχνικές Εφαρμογής Πλέγματος.....	- 14 -
3.2.1. Μέθοδος Προβολών	- 14 -
3.2.2. Μέθοδος με Αναγνώριση “σημαντικών” στιγμάτων (guide spots). -	16 -
3.2.3. Μέθοδος με χρήση Χαρτών απόστασης (distance maps)	- 16 -
3.2.4. Προσέγγιση “Ορειβασίας σε Λόφο” (Hill-Climbing Approach).....	- 17 -
3.2.5. Μπεϊσιανό ταίριασμα πλέγματος (Bayesian Grid Matching)	- 18 -
Μεθοδολογία Αυτόματης Εφαρμογής Πλέγματος	- 21 -

4.1.	Η Μεθοδολογία Αυτόματης Εφαρμογής Πλέγματος	- 21 -
4.1.1.	Αρχικά στάδια προεπεξεργασίας της εικόνας DNA	- 23 -
4.1.2.	Εκτίμηση της γωνίας περιστροφής και της ακτίνας των στιγμάτων -	24 -
4.1.3.	Αναζήτηση υποεικόνων (subarrays) και εύρεση των ορίων τους....	- 27 -
4.1.4.	Αναγνώριση των στιγμάτων και εύρεση των κέντρων τους	- 29 -
4.1.5.	Υπολογισμός των τετραγωνικών πυραμίδων.....	- 33 -
4.1.6.	Διαδικασία Εφαρμογής του Πλέγματος (gridding process).....	- 34 -
4.1.7.	Αποθήκευση των Δεδομένων του Πλέγματος	- 36 -
4.1.8.	Διαδικασία βελτίωσης του πλέγματος (Refinement procedure).....	- 37 -
	Αποτελέσματα Εφαρμογής Πλέγματος.....	- 41 -
5.1.	Αποτελέσματα Εφαρμογής Πλέγματος σε ιδανικές εικόνες DNA που έχουν παραμορφωθεί από γνωστούς μετασχηματισμούς	- 41 -
5.1.1.	Αφφινικός Μετασχηματισμός	- 42 -
5.1.2.	Ημιτονοειδής Μετασχηματισμός	- 44 -
5.2.	Αποτελέσματα Εφαρμογής Πλέγματος σε πραγματικές εικόνες DNA (άγνωστος μετασχηματισμός).....	- 45 -
5.3.	Συμπεράσματα για τη Προτεινόμενη Μεθοδολογία Αυτόματης Εφαρμογής Πλέγματος	- 49 -
5.4.	Σύγκριση Προτεινόμενης Αυτόματης Μεθοδολογίας με Μέθοδο Προβολών	- 50 -
	Αξιολόγηση	- 57 -
6.1.	Αξιολόγηση της Προτεινόμενης Αυτόματης Μεθοδολογίας	- 57 -
6.2.	Μελλοντικές Βελτιώσεις και Επεκτάσεις	- 57 -
	Βιβλιογραφία	- 59 -

Εικόνες – Σχήματα – Πίνακες

Σχήμα 2.1.: Δομή του DNA.....	- 6 -
Σχήμα 2.2.: Επισήμανση του mRNA.....	- 8 -
Σχήμα 2.3.: Ιδανική εικόνα DNA	- 9 -
Σχήμα 2.4.: Διαδικασία επεξεργασίας εικόνων DNA.....	- 10 -
Σχήμα 2.5.: Πραγματική εικόνα DNA	- 11 -
Σχήμα 4.1.: Μπλοκ διάγραμμα της προτεινόμενης μεθοδολογίας.....	- 22 -
Σχήμα 4.2.α.: Εικόνα DNA με θόρυβο	- 23 -
Σχήμα 4.2.β.: Αποτέλεσμα αποθορυβοποίησης	- 23 -
Σχήμα 4.3.: Εικόνα DNA με θόρυβο λόγω εσφαλμένης υβριδοποίησης.....	- 25 -
Σχήμα 4.4.: Εκτίμηση της διαμέτρου ενός στίγματος.....	- 25 -
Σχήμα 4.5.α.: Βήμα N = 1 αναζήτησης γωνίας περιστροφής ($\sigma = 44.6499$)	- 26 -
Σχήμα 4.5.β.: Βήμα N = 3 αναζήτησης γωνίας περιστροφής ($\sigma = 41.0172$)	- 26 -
Σχήμα 4.5.γ.: Βήμα N = 5 αναζήτησης γωνίας περιστροφής ($\sigma = 1.8771$).....	- 26 -
Σχήμα 4.5.δ.: Βήμα N = 10 αναζήτησης γωνίας περιστροφής ($\sigma = 19.8742$).....	- 26 -
Σχήμα 4.6.α.: Βήμα N=1 αναζήτησης υποεικόνων	- 28 -
Σχήμα 4.6.β.: Βήμα N=2 αναζήτησης υποεικόνων	- 28 -
Σχήμα 4.6.γ.: Βήμα N=3 αναζήτησης υποεικόνων.....	- 28 -
Σχήμα 4.6.δ.: Βήμα N=4 αναζήτησης υποεικόνων	- 28 -
Σχήμα 4.7.α.: Εικόνα εισόδου για ανίχνευση υποεικόνων	- 28 -
Σχήμα 4.7.β.: Επιτυχής ανίχνευση των υποεικόνων.....	- 28 -
Σχήμα 4.8.: Διαδικασία ταιριάσματος προτύπων	- 29 -
Σχήμα 4.9.α.: Πρότυπο προς αναγνώριση	- 31 -
Σχήμα 4.9.β.: Εικόνα αναζήτησης	- 31 -
Σχήμα 4.10.α.: Ταίριασμα εικόνας με χρήση της μετρικής “mxd”	- 31 -
Σχήμα 4.10.β.: Ταίριασμα εικόνας με χρήση της μετρικής “sad”	- 31 -
Σχήμα 4.10.γ.: Ταίριασμα εικόνας με χρήση της μετρικής “ssd”	- 31 -
Σχήμα 4.10.δ.: Ταίριασμα εικόνας με χρήση της μετρικής “ccr”	- 31 -
Σχήμα 4.10.ε.: Ταίριασμα εικόνας με χρήση της μετρικής “ncc”	- 31 -
Σχήμα 4.10.ζ.: Ταίριασμα εικόνας με χρήση της μετρικής “ccf”	- 31 -

Σχήμα 4.11.α.: Εικόνα εισόδου για ανίχνευση των κέντρων των στιγμάτων	- 32 -
Σχήμα 4.11.β.: Τοπικά μέγιστα με τη κατάλληλη ακτίνα αναζήτησης	- 32 -
Σχήμα 4.11.γ.: Τοπικά μέγιστα με πολύ μικρή ακτίνα αναζήτησης	- 32 -
Σχήμα 4.11.δ.: Τοπικά μέγιστα με πολύ μεγάλη ακτίνα αναζήτησης	- 32 -
Σχήμα 4.12.α.: Εικόνα εισόδου	- 34 -
Σχήμα 4.12.β.: Πυραμίδες με κατάλληλη πλευρά	- 34 -
Σχήμα 4.12.γ.: Πυραμίδες με πολύ μικρή πλευρά	- 34 -
Σχήμα 4.12.δ.: Πυραμίδες με πολύ μεγάλη πλευρά	- 34 -
Σχήμα 4.13.α.: Οριζόντιο “μέτωπο” εικόνας πυραμίδων	- 35 -
Σχήμα 4.13.β.: Κατακόρυφο “μέτωπο” εικόνας πυραμίδων	- 35 -
Σχήμα 4.14.α.: Πίνακας αποθήκευσης $xGrid(i,jCell)$	- 36 -
Σχήμα 4.14.β.: Πίνακας αποθήκευσης $yGrid(iCell,j)$	- 36 -
Σχήμα 4.15.α.: Αποτέλεσμα πριν την διαδικασία βελτίωσης	- 40 -
Σχήμα 4.15.β.: Αποτέλεσμα μετά την διαδικασία βελτίωσης	- 40 -
Σχήμα 5.1.α.: Ιδανική εικόνα εισόδου DNA (χωρίς μετασηματισμό)	- 42 -
Σχήμα 5.1.β.: Αποτέλεσμα εφαρμογής πλέγματος σε ιδανική εικόνα DNA	- 42 -
Σχήμα 5.2.α.: Ιδανική εικόνα DNA παραμορφωμένη από Αφφινικό-1	- 43 -
Σχήμα 5.2.β.: Αποτέλεσμα εφαρμογής πλέγματος στην εικόνα με Αφφινικό-1	- 43 -
Σχήμα 5.3.α.: Ιδανική εικόνα DNA παραμορφωμένη από Αφφινικό-2	- 43 -
Σχήμα 5.3.β.: Αποτέλεσμα εφαρμογής πλέγματος στην εικόνα με Αφφινικό-2	- 43 -
Σχήμα 5.4.α.: Ιδανική εικόνα DNA παραμορφωμένη από Ημιτονοειδή-1	- 44 -
Σχήμα 5.4.β.: Αποτέλεσμα εφαρμογής πλέγματος στην εικόνα με Ημιτονοειδή -1-	44 -
-	
Σχήμα 5.5.α.: Ιδανική εικόνα DNA παραμορφωμένη από Ημιτονοειδή-2	- 45 -
Σχήμα 5.5.β.: Αποτέλεσμα εφαρμογής πλέγματος στην εικόνα με Ημιτονοειδή -2-	45 -
-	
Σχήμα 5.6.α.: Υποεικόνα (1,2) της πραγματικής εικόνας DNA	- 46 -
Σχήμα 5.6.β.: Αποτέλεσμα εφαρμογής πλέγματος στην υποεικόνα (1,2)	- 46 -
Σχήμα 5.6.γ.: Αποτέλεσμα εφαρμογής πλέγματος στην υποεικόνα (1,2) σε μεγέθυνση	- 46 -
Σχήμα 5.7.α.: Υποεικόνα (2,2) της πραγματικής εικόνας DNA	- 47 -

Σχήμα 5.7.β.: Αποτέλεσμα εφαρμογής πλέγματος στην υποεικόνα (2,2).....	- 47 -
Σχήμα 5.7.γ.: Αποτέλεσμα εφαρμογής πλέγματος στην υποεικόνα (2,2) σε μεγέθυνση	- 47 -
Σχήμα 5.9.: Αποτέλεσμα εφαρμογής πλέγματος σε πραγματική εικόνα DNA με 4 υποεικόνες.....	- 48 -
Πίνακας 5.1. Αξιολόγηση αποτελεσμάτων της Προτεινόμενης Αυτόματης Μεθοδολογίας	- 49 -
Σχήμα 5.10.α.: Αποτέλεσμα Μεθόδου Προβολών σε ιδανική εικόνα DNA	- 50 -
Σχήμα 5.10.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας σε ιδανική εικόνα DNA ..	- 50 -
Σχήμα 5.11.α.: Αποτέλεσμα Μεθόδου Προβολών στην εικόνα με Αφφινικό-1....	- 51 -
Σχήμα 5.11.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην εικόνα με Αφφινικό-1.....	- 51 -
Σχήμα 5.12.α.: Αποτέλεσμα Μεθόδου Προβολών στην εικόνα με Αφφινικό-2....	- 51 -
Σχήμα 5.12.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην εικόνα με Αφφινικό-2.....	- 51 -
Σχήμα 5.13.α.: Αποτέλεσμα Μεθόδου Προβολών στην εικόνα με Ημιτονοειδή-1-	- 52 -
Σχήμα 5.13.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην εικόνα με Ημιτονοειδή-1	- 52 -
Σχήμα 5.14.α.: Αποτέλεσμα Μεθόδου Προβολών στην εικόνα με Ημιτονοειδή-2-	- 52 -
Σχήμα 5.14.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην εικόνα με Ημιτονοειδή-2	- 52 -
Σχήμα 5.15.α.: Αποτέλεσμα Μεθόδου Προβολών στην υποεικόνα (1,2)	- 53 -
Σχήμα 5.15.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην υποεικόνα (1,2)-	- 53 -
Σχήμα 5.16.α.: Αποτέλεσμα Μεθόδου Προβολών στην υποεικόνα (2,2)	- 53 -
Σχήμα 5.16.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην υποεικόνα (2,2)-	- 53 -
Σχήμα 5.17.α.: Αποτέλεσμα Μεθόδου Προβολών στην πραγματική εικόνα DNA με υποεικόνες.....	- 54 -
Σχήμα 5.17.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην πραγματική εικόνα DNA.....	- 54 -
Πίνακας 5.2. Αξιολόγηση αποτελεσμάτων της Μεθόδου των Προβολών	- 55 -
Πίνακας 5.3. Συγκριτικός Πίνακας Αποτελεσμάτων	- 55 -

Κεφάλαιο

1

Εισαγωγή

1.1. Σκοπός Διπλωματικής Εργασίας

Η αναζήτηση καινούριων, αποτελεσματικότερων θεραπειών για διάφορες ασθένειες, όπως για παράδειγμα ο καρκίνος, αλλά και η ανάγκη των ερευνητών να μάθουν περισσότερα για τη δομή και λειτουργία του ανθρώπινου οργανισμού είναι κάποιοι από τους λόγους που ώθησαν στην ανάπτυξη της τεχνολογίας των εικόνων DNA, γνωστή στην διεθνή βιβλιογραφία ως DNA microarrays technology.

Πολλές από τις έρευνες που χρησιμοποιούν αυτή την τεχνολογία, γίνονται για τον καρκίνο. Στο παρελθόν, οι επιστήμονες συνήθιζαν να χαρακτηρίζουν τα διάφορα είδη καρκίνου με βάση τα όργανα, στα οποία αναπτύσσονταν οι όγκοι. Με τη βοήθεια, ωστόσο, της τεχνολογίας των εικόνων DNA, οι επιστήμονες θα είναι σε θέση να ταξινομήσουν περαιτέρω αυτούς τους τύπους καρκίνου με βάση τα πρότυπα της γονιδιακής δραστηριότητας στα κύτταρα του όγκου και θα είναι σε θέση να σχεδιάσουν στρατηγικές θεραπείας που απευθύνονται άμεσα σε κάθε συγκεκριμένο τύπο καρκίνου.

Η τεχνολογία αυτή των εικόνων DNA, βοηθά τους ερευνητές να μελετήσουν περισσότερο διάφορες ασθένειες όπως καρδιοπάθειες, ψυχικές διαταραχές και λοιμώξεις. Εξετάζοντας τις διαφορές στη δραστηριότητα γονιδίων μεταξύ θεραπευμένων και μη θεραπευμένων κυττάρων του όγκου - για παράδειγμα, εκείνα που είναι ακτινοβολημένα ή αυτά που έχουν υποστεί έλλειψη οξυγόνου - οι

επιστήμονες θα κατανοήσουν ακριβώς πώς διάφορες θεραπείες επηρεάζουν τους όγκους, ώστε να είναι σε θέση να αναπτύξουν πιο αποτελεσματικές θεραπείες.

Από τη φύση της κατασκευής των εικόνων αυτών, φαίνεται ότι τα δεδομένα που χρειάζονται οι ερευνητές είναι κρυμμένα στην πληροφορία της φωτεινότητας, του σχήματος αλλά και του μεγέθους των στιγμάτων που περιέχονται σε μια τέτοια εικόνα. Επομένως, η ανάλυση των δεδομένων αυτών, προϋποθέτει την ανάλυση της ίδιας της εικόνας, με τεχνικές ψηφιακής επεξεργασίας εικόνων.

Αν ακολουθήσει κανείς τα βήματα που απαιτούνται για την επεξεργασία αυτών των εικόνων, θα συμπεράνει την αναγκαιότητα για τη χρήση ενός αλγορίθμου εφαρμογής πλέγματος, ή όπως συνηθίζεται στην διεθνή βιβλιογραφία *gridding process*, στις εικόνες αυτές. Πολλές φορές, γνωρίζοντας το επιθυμητό αποτέλεσμα, είναι ο μόνος τρόπος για να ξεκαθαρίσει η διαδικασία η οποία πρέπει να ακολουθηθεί για να οδηγηθούμε στον τελικό στόχο.

Ξεκινώντας λοιπόν από το τέλος, πρέπει να έχουμε την πληροφορία της φωτεινότητας των στιγμάτων, την ακτίνα και το μέγεθος τους. Σε αρκετές περιπτώσεις, ιδιαίτερα για αξιολόγηση της ποιότητας των δεδομένων μας (βλ. παράγραφο 2.4.3), σημαντικό ρόλο παίζει και το περίγραμμα των στιγμάτων αυτών. Δίνοντας μια ιδέα για το μέγεθος των εικόνων DNA, ώστε να έχει ο αναγνώστης μια ξεκάθαρη εικόνα για το πλήθος των στιγμάτων, οι εικόνες αυτές φτάνουν τάξεις μεγέθους δεκάδων χιλιάδων εικονοστοιχείων (*pixels*), γεγονός που δυσκολεύει υπερβολικά την ταυτόχρονη επεξεργασία αυτών των στιγμάτων.

Σε αυτό το σημείο, φαίνεται καθαρά η ανάγκη ανάπτυξης μεθοδολογιών για αυτόματη εφαρμογή πλέγματος, προκειμένου να προχωρήσουμε στα επόμενα βήματα επεξεργασίας, όπως είναι η κατάτμηση, η οποία δίνει άμεσα την πληροφορία για το μέγεθος και το περίγραμμα των στιγμάτων.

Το πλέγμα είναι ένας πίνακας ή γενικότερα ένα σύνολο διατεταγμένων ζευγών αριθμών (συντεταγμένων, συνήθως) στους οποίους είναι αποθηκευμένη η πληροφορία για το που βρίσκεται το κάθε στίγμα στην εικόνα DNA και ποιες οι ακριβείς συντεταγμένες του σε αυτήν. Με αυτό τον τρόπο, είναι προφανής η ανάγκη χρήσης του για απομόνωση του εκάστοτε στίγματος, για άμεση αναφορά σε αυτό, επιλέγοντας τη γραμμή και τη στήλη στις οποίες ανήκει και ίσως για μεγέθυνση της περιοχής του για περαιτέρω επεξεργασία της. Αυτός ακριβώς είναι και ο σκοπός της παρούσας διπλωματικής εργασίας. Η υλοποίηση ενός τέτοιου αλγορίθμου για αυτόματη εφαρμογή πλέγματος.

1.2. Διάρθρωση Διπλωματικής Εργασίας

Στο υπόλοιπο μέρος της διπλωματικής εργασίας, γίνεται μια εισαγωγή στις σημαντικότερες έννοιες που χρησιμοποιούνται, παρουσιάζονται διάφορες τεχνικές

εφαρμογής πλέγματος που ήδη χρησιμοποιούνται, εξηγείται βήμα προς βήμα η προτεινόμενη μεθοδολογία και παρουσιάζονται κάποια ενδεικτικά αποτελέσματα συγκριτικά με τις προϋπάρχουσες μεθόδους.

Αναλυτικότερα, στο δεύτερο κεφάλαιο, γίνεται μια εισαγωγή στα γονίδια, το DNA και τις πρωτεΐνες, έννοιες σημαντικότερες ώστε ο αναγνώστης να είναι σε θέση να κατανοήσει την αρχή λειτουργίας της τεχνολογίας των εικόνων DNA. Στο δεύτερο μέρος του κεφαλαίου, δίνεται μια σύντομη περιγραφή της τεχνολογίας των εικόνων DNA, την οποία ακολουθεί μια λεπτομερέστερη περιγραφή των επιμέρους βημάτων που λαμβάνουν χώρα από την στιγμή που συλλέγονται τα δείγματα μέχρι και τα τελευταία στάδια ψηφιοποίησης των εικόνων. Το τελευταίο μέρος, περιγράφει τα στάδια της διαδικασίας ανάγνωσης αυτών των εικόνων καθώς και τα πρότυπα με τα οποία είναι αναγκαία να συμμορφώνονται οι εικόνες αυτές για να είναι αναγνώσιμες και συγκρίσιμες από εργαστήρια ανά το παγκόσμιο.

Το σύνολο των μεθοδολογιών εφαρμογής πλέγματος που παρουσιάστηκαν μέχρι τώρα είναι η κεντρική ιδέα του τρίτου κεφαλαίου. Στο πρώτο μέρος του κεφαλαίου αυτού, παρουσιάζεται επιγραμματικά η γενική διαδικασία εφαρμογής πλέγματος σε εικόνες DNA. Στο δεύτερο μέρος του τρίτου κεφαλαίου, παρουσιάζονται λεπτομερώς οι μεθοδολογίες που ήδη χρησιμοποιούνται από τους ερευνητές. Μαζί με την πλήρη περιγραφή της εκάστοτε μεθοδολογίας, δίνονται και κάποια μειονεκτήματα που ίσως παρουσιάζονται σε συγκεκριμένες συνθήκες.

Στο τέταρτο κεφάλαιο, δίνεται μια λεπτομερής περιγραφή της αυτόματης μεθόδου που προτείνεται στα πλαίσια της παρούσης διπλωματικής εργασίας. Ο αναγνώστης σε αυτό το κεφάλαιο οδηγείται βήμα προς βήμα στα στάδια υλοποίησης της προτεινόμενης μεθόδου, με στόχο την πλήρη κατανόηση της διαδικασίας που ακολουθήθηκε. Εξηγείται πως έγινε η αρχική σύλληψη της ιδέας και γιατί υποστηρίζεται ότι η συγκεκριμένη μέθοδος μπορεί να επιτύχει σε συνθήκες κάτω από τις οποίες οι προηγούμενες μέθοδοι απέτυχαν. Επιπρόσθετα, αναφέρεται ο τρόπος επιλογής σημαντικών, για την μέθοδο, παραμέτρων, αλλά και των διαφόρων κατωφλίων που χρησιμοποιήθηκαν, δίνοντας τεκμηριωμένες εξηγήσεις με πίνακες, γραφικές παραστάσεις και διαγράμματα.

Για απόδειξη των ισχυρισμών περί υπεροχής της προτεινόμενης μεθόδου, παρουσιάζονται αποτελέσματα σε εικόνες με εφαρμογή της μεθόδου σε δυσμενείς συνθήκες. Επιπλέον, πραγματοποιείται σύγκριση αποτελεσμάτων της προτεινόμενης αυτόματης εφαρμογής πλέγματος με τα αντίστοιχα αποτελέσματα από τη μέθοδο προβολών, μια ενδεικτική μέθοδο που χρησιμοποιείται ευρέως από τους ερευνητές.

Τέλος, στο έκτο κεφάλαιο, παρουσιάζονται κάποια συμπεράσματα για την προτεινόμενη μέθοδο καθώς και η αξιολόγηση της σε σχέση με τα αποτελέσματα που προέκυψαν από την εφαρμογή της.

Εικόνες DNA (DNA microarrays)

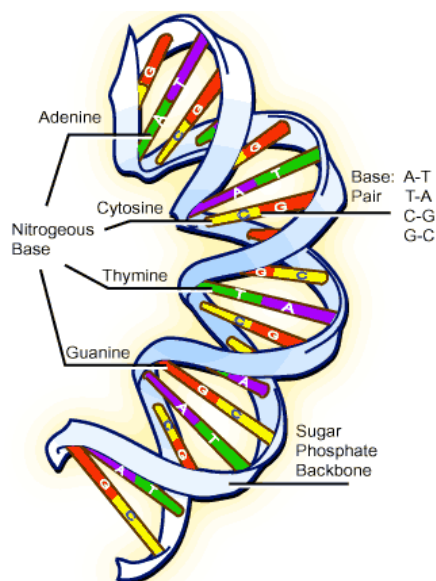
2.1. Εισαγωγή στα γονίδια, DNA και πρωτεΐνες

Πριν προχωρήσουμε στην περιγραφή των εικόνων DNA, θα πρέπει πρώτα να γίνει μια μικρή εισαγωγή για να αναφερθούμε στα γονίδια, τις πρωτεΐνες και την δημιουργία τους καθώς και στο DNA (δεσοξυριβοζονουκλεϊνικό οξύ).

Οι περισσότεροι ευκαρυωτικοί οργανισμοί (οι οργανισμοί των οποίων τα κύτταρα είναι οργανωμένα σε πολύπλοκες δομές εγκλεισμένες μέσα σε μεμβράνες), όπως για παράδειγμα ο άνθρωπος, αποτελούνται από δισεκατομμύρια κύτταρα. Σχεδόν όλα αυτά τα κύτταρα περιέχουν, στο πυρήνα τους, ολόκληρο το γονιδίωμα για το συγκεκριμένο οργανισμό. Αυτό το γονιδίωμα περιέχει όλη την κληρονομική πληροφορία του οργανισμού στη μορφή του δεσοξυριβοζονουκλεϊνικού οξέος (DNA) το οποίο κωδικοποιεί ένα ολοκληρωμένο πρότυπο για όλες τις δραστηριότητες και τις δομές του συγκεκριμένου οργανισμού [2].

Στο ανθρώπινο σώμα, το γονιδίωμα αποτελείται από 23 ζευγάρια χρωμοσωμάτων. Το ένα, από καθένα από αυτά τα ζευγάρια, κληρονομείται από την μητέρα και το άλλο από τον πατέρα. Το κάθε χρωμόσωμα αποτελείται από αλυσίδες DNA η οποίες με την σειρά τους αποτελούνται από δύο πολυμερή (μεγάλα μόρια από επαναλαμβανόμενες υποομάδες) τα οποία συντελούνται από τα νουκλεοτίδια. Τα μόρια αυτά είναι δεμένα μεταξύ τους σε μορφή διπλής έλικας. Κάθε νουκλεοτίδιο

αποτελείται από δεσοξυριβόζη, μια φωσφορική ομάδα και μια από τις αζωτούχες βάσεις, κυανίνη, αδενίνη, θυμίνη και κυτοσίνη. Συνήθως, αναφερόμαστε σε αυτές τις βάσεις, με τα αρχικά τους, G, A, T και C, αντίστοιχα. Ουσιαστικά, εκεί βρίσκεται κωδικοποιημένη η κληρονομική γενετική πληροφορία. Αξίζει να σημειωθεί ότι μόνο ένα από τα δύο σκέλη της διπλής έλικας DNA αρκεί για να περιγράψει αυτή την πληροφορία. Αυτό οφείλεται στη συμπληρωματική αντιστοίχιση των ζευγαριών κατά την οποία το A στο ένα σκέλος πάντοτε δένει με ένα G στο άλλο σκέλος καθώς ένα T δένει πάντοτε με ένα C (βλ. Σχήμα 2.1.) [2].



Σχήμα 2.1.: Δομή του DNA

Τα γονίδια είναι τμήματα της δομής του DNA που περιγράφεται παραπάνω. Ουσιαστικά, ένα γονίδιο είναι ένα τμήμα του DNA που καθορίζει ένα ενιαίο γνώρισμα με την κωδικοποίηση ενός συγκεκριμένου πρότυπου, από τα περίπου 27.000 που υπάρχουν στον άνθρωπο. Πρακτικά, ένα γονίδιο είναι μια εντοπίσιμη περιοχή της ακολουθίας του γονιδιώματος, που αντιστοιχεί σε μια μονάδα της γενετικής κληρονομιάς [2].

Ο κύριος σκοπός των γονιδίων είναι η λειτουργία τους ως πρότυπο για τη δημιουργία των πρωτεϊνών. Οι Πρωτεΐνες είναι τα δομικά στοιχεία του οργανισμού, αποτελούνται από αμινοξέα και είναι υπεύθυνα για την δομή και δραστηριότητα ενός οργανισμού σε επίπεδο κυττάρου. Μολονότι τα φυσικά αμινοξέα από τα οποία δομούνται οι πρωτεΐνες είναι μόνο 20, οι διάφοροι μεταξύ τους συνδυασμοί παρέχουν ένα τεράστιο αριθμό πρωτεϊνικών μορίων. Τα αμινοξέα είναι οργανικές ενώσεις οι οποίες περιέχουν στο μόριο τους μια καρβοξυλομάδα και μια αμινομάδα συνδεδεμένες στο ίδιο άτομο άνθρακα. Ο δεσμός μεταξύ δύο αμινοξέων ονομάζεται πεπτιδικός και δημιουργείται ύστερα από την αντίδραση της αμινομάδας ενός αμινοξέος με την καρβοξυλομάδα ενός άλλου. Με αυτό τον τρόπο σχηματίζονται οι πολυπεπτιδικές αλυσίδες, οι οποίες στη συνέχεια συνδέονται κατάλληλα προκειμένου να δημιουργήσουν τα πρωτεϊνικά μόρια [13].

Η πληροφορία που περιέχεται στο γονίδιο μεταγράφεται σε έναν αγγελιοφόρο σκέλος, το ριβοζονουκλεϊνικό οξύ (mRNA). Αυτή η διαδικασία γίνεται από ένα ένζυμο που ονομάζεται RNA πολυμεράση. Μετά την μεταγραφή αυτή του μορίου mRNA, αυτό αφήνει τον πυρήνα του κυττάρου όπου μεταγράφεται σε μια πρωτεΐνη. Η διαδικασία αυτή ονομάζεται “μετάφραση”. Αυτό γίνεται από τα ριβοσώματα, που διαβάζουν τον κώδικα που μεταφέρεται από το μόριο mRNA, από τον πυρήνα των κυττάρων, και δημιουργούν τις πρωτεΐνες. Αυτή η διαδικασία της

μετάφρασης ενός γονιδίου σε ένα λειτουργικό προϊόν είναι γνωστή ως “ γονιδιακή έκφραση” (gene expression) [2].

2.2. Περιγραφή της τεχνολογίας των εικόνων DNA (DNA microarrays technology)

Παρά το γεγονός ότι όλα τα κύτταρα στο ανθρώπινο σώμα περιέχουν πανομοιότυπο γενετικό υλικό, τα ίδια γονίδια δεν είναι ενεργά σε κάθε κύτταρο. Μελετώντας ποια γονίδια είναι ενεργά και πια ανενεργά σε διαφορετικούς τύπους κυττάρων, δίνεται η ευκαιρία στους επιστήμονες να κατανοήσουν τον τρόπο με τον οποίο τα κύτταρα αυτά λειτουργούν κανονικά καθώς και πως επηρεάζονται όταν διάφορα γονίδια δεν λειτουργούν σωστά.

Οι εικόνες DNA, είναι μια υψηλής ανάλυσης τεχνολογία η οποία χρησιμοποιείται για τη μέτρηση των επιπέδων έκφρασης χιλιάδων γονιδίων ή σε κάποιες περιπτώσεις όλων των γονιδίων, σε ένα γονιδίωμα, ταυτόχρονα. Η αρχή λειτουργίας της τεχνολογίας των πλείστων εικόνων DNA, είναι η εκμετάλλευση της συμπληρωματικής αντιστοίχισης των ζευγαριών (βλ. παράγραφο 2.1. για περαιτέρω ανάλυση) για μέτρηση του πλήθους των διαφορετικών τύπων μορίων mRNA σε ένα κύτταρο, μετρώντας εμμέσως με αυτό το τρόπο τα επίπεδα έκφρασης των γονιδίων που είναι υπεύθυνα για την σύνθεση των συγκεκριμένων μορίων mRNA [5].

Τα στίγματα σε αυτές τις εικόνες περιέχουν μονοσκελή ολιγονουκλεοτίδια DNA αποκαλούμενα ως καθετήρες (probes). Κάθε μια από αυτές τις κηλίδες θα περιέχει μια αλληλουχία DNA που είναι συμπληρωματική για το συγκεκριμένο μόριο mRNA που αντιστοιχεί στο στοχευόμενο γονίδιο. Ένα μόριο mRNA που είναι συμπληρωματικό ως προς το καθετήρα, θα πρέπει να υβριδοποιηθεί με τον καθετήρα και να αποτελέσει έναν ισχυρό δεσμό μεταξύ των mRNA - DNA. Τα μόρια mRNA θα έχουν προηγουμένως επισημανθεί με τις φθορίζουσες χρωστικές ουσίες, επομένως το ποσό του υβριδισμού μπορεί να μετρηθεί με το επίπεδο του φθορισμού της χρωστικής ουσίας, η οποία εξετάζεται με ένα σαρωτή (scanner) [2].

Ο σαρωτής αυτός στη συνέχεια, βγάζει στην έξοδο του ένα αρχείο για κάθε πίνακα, το οποίο περιέχει τα στοιχεία σχετικά με αυτό τον πίνακα, όπως το επίπεδο του φθορισμού του κάθε σημείου και το επίπεδο του θορύβου του περιβάλλοντος. Αυτά είναι τα αρχεία, που στη συνέχεια, αναλύονται υπολογιστικά. Θεωρητικά, ένα σημείο με φωτεινότερο φθορισμό σημαίνει ότι περισσότερο mRNA έχει υβριδοποιηθεί, το οποίο με τη σειρά του συνεπάγεται ότι περισσότερα mRNA ήταν παρούσα στο δείγμα που προέρχονται από το αρχικό κύτταρο και ότι το γονίδιο εκπροσωπούμενο από αυτό το σημείο βρίσκεται σε ένα υψηλότερο επίπεδο γονιδιακής έκφρασης (gene expression) [2].

Οι πιο ευρέως χρησιμοποιούμενοι τύποι εικόνων DNA, μπορούν να ταξινομηθούν σε δύο κατηγορίες, τις cDNA συστοιχίες και τις συστοιχίες ολιγονουκλεοτιδίων.

2.3. Διαδικασία δημιουργίας των εικόνων DNA

Γενικά, η διαδικασία από τη συλλογή των δειγμάτων μέχρι και τη ψηφιοποίηση της εικόνας DNA αποτελείται από τα ακόλουθα πέντε διαφορετικά στάδια.

2.3.1. Συλλογή Δεδομένων

Επιλέγονται δύο τύποι δειγμάτων, ένα δείγμα αναφοράς, για παράδειγμα ένα υγιές κύτταρο και το υπό εξέταση δείγμα, για παράδειγμα ένα καρκινικό κύτταρο. Τα δείγματα τοποθετούνται με ρομποτικά μηχανήματα τα οποία στοιχίζουν εκατοντάδες ή χιλιάδες γονιδιακές αλληλουχίες σε μια ενιαία πλάκα μικροσκοπίου.

2.3.2. Απομόνωση του αγγελιαφόρου RNA (mRNA)

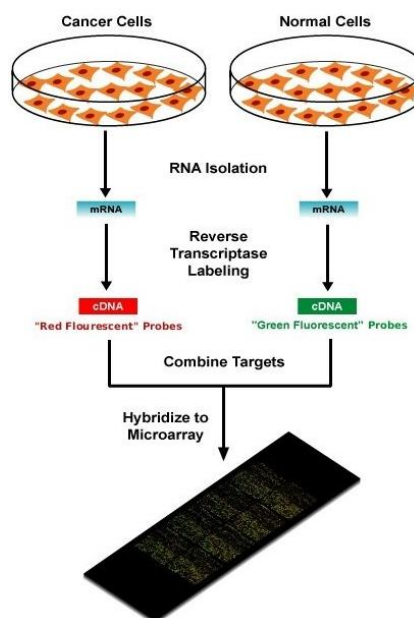
Σε κάθε ένα από τα δύο κύτταρα ο ερευνητής απομονώνει το αγγελιαφόρο RNA (mRNA). Προκειμένου να προσδιοριστούν ποια γονίδια είναι ενεργοποιημένα και πια όχι σε ένα δεδομένο κύτταρο, ο ερευνητής πρέπει πρώτα να συλλέξει τα μόρια αγγελιοφόρου RNA (mRNA) που είναι παρόντα στο συγκεκριμένο κελί. Όταν ένα γονίδιο ενεργοποιείται, το κινητό μηχάνημα αρχίζει να αντιγράφει ορισμένα τμήματα του εν λόγω γονιδίου.

2.3.3. Επισήμανση του mRNA

Ο ερευνητής, επισημαίνει κάθε μόριο mRNA, χρησιμοποιώντας ένα ένζυμο ανάστροφης μεταγράψης (reverse transcriptase - RT), που δημιουργεί ένα συμπληρωματικό cDNA του mRNA. Κατά τη διάρκεια αυτής της διαδικασίας, τα φθορίζοντα νουκλεοτίδια προσκολλούνται στο cDNA. Τόσο τα καρκινικά όσο και τα κανονικά δείγματα επισημαίνονται με διαφορετικές φθορίζουσες χρωστικές ουσίες. Το υγιές δείγμα βάφεται με πράσινη χρωστική ουσία, ενώ το καρκινικό δείγμα βάφεται με κόκκινη χρωστική ουσία (βλ. Σχήμα 2.2.).

2.3.4. Υβριδοποίηση

Η εικόνα αποτελείται από πολλά (10.000 ή και περισσότερα) στίγματα, που το καθένα περιέχει ένα ολιγονουκλεοτίδιο που αντιστοιχεί σε 1 γονίδιο. Στην πραγματικότητα, κάθε ένα στίγμα, περιέχει εκατομμύρια τέτοια



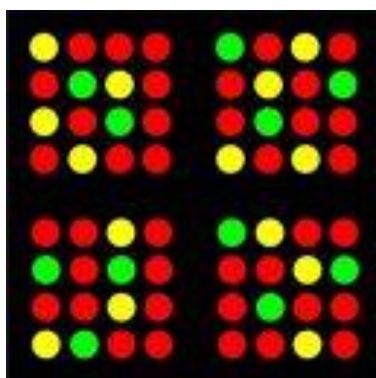
Σχήμα 2.2.: Επισήμανση του mRNA

ολιγονουκλεοτίδια. Επομένως, υπάρχει περιθώριο για πολλά cDNA να δεσμευτούν πάνω τους. Όσο περισσότερα cDNA στο δείγμα, τόσο περισσότερο τείνει να συνδεθεί σε εκείνο το σημείο. Σε αυτό το στάδιο, ο ερευνητής τοποθετεί την ετικέτα cDNA στη διαφάνεια των εικόνων DNA. Στη συνέχεια, τα επισημασμένα cDNAs, που αντιπροσωπεύουν mRNAs στο κελί, θα βρουν το συγκεκριμένο σημείο τους και θα υβριδοποιηθούν - ή θα δεσμευτούν - με τα συνθετικά συμπληρωματικά DNAs που βρίσκονται στη διαφάνεια της εικόνας, αφήνοντας την φθορίζουσα ετικέτα τους.

2.3.5. Καθαρισμός Εικόνων και Μέτρηση Φωτεινότητας

Η επιφάνεια πλένεται για να απομακρυνθούν αδέσμευτα cDNAs. Ο ερευνητής στη συνέχεια, πρέπει να χρησιμοποιήσει ένα ειδικό σαρωτή για τη μέτρηση της έντασης φθορισμού για κάθε στίγμα/περιοχή στη διαφάνεια των εικόνων. Η εικόνα έχει σαρωθεί με 2 χρώματα, ένα για ανίχνευση της κόκκινης βαφής και ένα για την πράσινη βαφή. Οι δύο σαρώσεις στη συνέχεια θα τοποθετηθούν η μια πάνω στην άλλη.

Σε περίπτωση που ένα συγκεκριμένο γονίδιο, είναι πολύ ενεργό, παράγει πολλά μόρια του αγγελιοφόρου RNA (mRNA), επομένως και περισσότερα επισημασμένα cDNAs, τα οποία υβριδοποιούν το DNA στη διαφάνεια της εικόνας και δημιουργούν μια πολύ φωτεινή περιοχή φθορισμού. Τα γονίδια που είναι λιγότερο ενεργά, παράγουν λιγότερα mRNAs και κατά συνέπεια, λιγότερα επισημασμένα cDNAs, γεγονός που οδηγεί σε διαφορετικού φθορισμού περιοχές. Αν δεν υπάρχει φθορισμός, κανένα από τα μόρια αγγελιοφόρου δεν έχει υβριδοποιηθεί στο DNA, συμπεραίνοντας ότι το γονίδιο είναι ανενεργό.



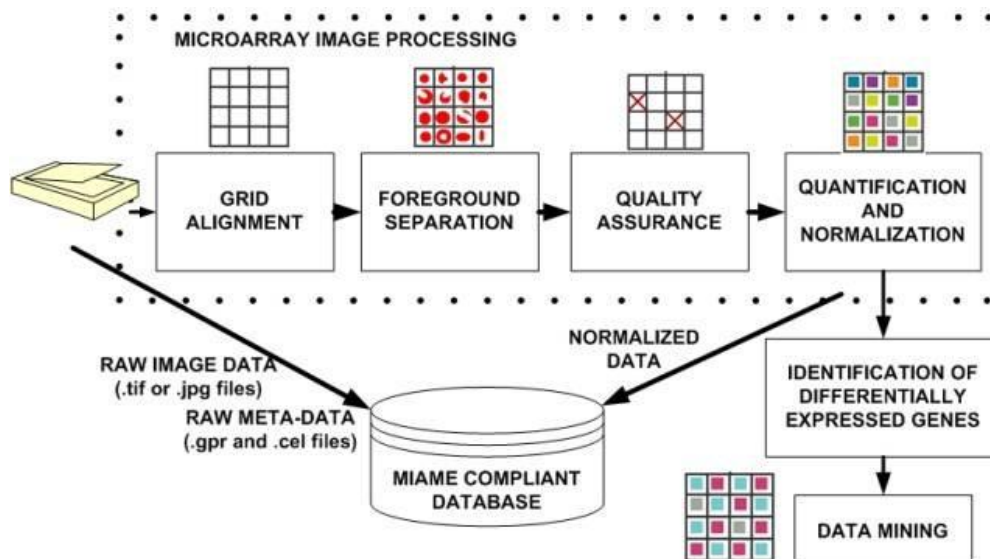
Σχήμα 2.3.: Ιδανική εικόνα DNA

Οι ερευνητές χρησιμοποιούν συχνά αυτή την τεχνική για να εξετάσουν τη δραστηριότητα των διαφόρων γονιδίων σε διαφορετικές χρονικές στιγμές. Όταν υβριδοποιούνται ταυτόχρονα καρκινικά δείγματα (κόκκινη χρωστική ουσία) και κανονικά δείγματα (πράσινη χρωστική ουσία), θα ανταγωνιστούν για τα συνθετικά συμπληρωματικά DNAs στη διαφάνεια της εικόνας (βλ. Σχήμα 2.3.). Επομένως, εάν ένα στίγμα είναι κόκκινο, σημαίνει ότι το συγκεκριμένο γονίδιο είναι καρκινικό. Αν ένα στίγμα είναι πράσινο, σημαίνει ότι αυτό το γονίδιο είναι υγιές. Αν ένα στίγμα είναι κίτρινο, αυτό σημαίνει ότι το συγκεκριμένο γονίδιο είναι εξίσου εκτεθειμένο σε καρκινικά όσο και σε υγιή.

2.4. Διαδικασία Ανάγνωσης Εικόνων DNA

Ο στόχος της ανάλυση εικόνων DNA είναι η εξαγωγή αποτελεσμάτων για την ένταση από κάθε σημείο της εικόνας που αντιπροσωπεύουν τα επίπεδα έκφρασης των γονιδίων και δυνατότητες εισαγωγής για περαιτέρω ανάλυση. Τα βιολογικά συμπεράσματα, στη συνέχεια, συντάσσονται με βάση τα αποτελέσματα από την εξόρυξη των δεδομένων και στατιστική ανάλυση όλων των εξαγόμενων χαρακτηριστικών.

Η ανάλυση των εικόνων DNA για την παραγωγή καταλόγων διαφορετικά εκφρασμένων γονιδίων, όπως φαίνεται στο Σχήμα 2.4, έχει πολλά βήματα τα οποία μπορεί να διαφέρουν ανάλογα με το είδος των δεδομένων που αποτελεί αντικείμενο της δοκιμασίας. Τα βήματα τα οποία περιλαμβάνει είναι: η ανάγνωση των ανεπεξέργαστων δεδομένων, η εφαρμογή πλέγματος για άμεση αναφορά σε στίγματα, ο διαχωρισμός τους από το υπόβαθρο (background), η αξιολόγηση της ποιότητας των δεδομένων, η απομάκρυνση των εσφαλμένων (χαμηλής ποιότητας και αξιοπιστίας) στιγμάτων για περαιτέρω ανάλυση, η κβαντοποίηση - κανονικοποίηση, η επεξεργασία των δεδομένων και ο υπολογισμός διαφορικής έκφρασης μέσω στατιστικής ανάλυσης. Τέλος, είναι πολύ σημαντικό τα αποτελέσματα των πειραμάτων να συμμορφώνονται σε κάποια διεθνή πρότυπα, ώστε να είναι αναγνώσιμα και συγκρίσιμα με αντίστοιχα άλλων εργαστηρίων [5].



Σχήμα 2.4.: Διαδικασία επεξεργασίας εικόνων DNA

2.4.1. Ανάγνωση Ανεπεξέργαστων Δεδομένων

Μόλις καθαριστεί η πλάκα της εικόνας DNA, διαβάζονται τα ανεπεξέργαστα δεδομένα, ως έχουν, τα οποία ουσιαστικά είναι η φωτεινότητα φθορισμού του εκάστοτε στίγματος πάνω στην πλάκα.

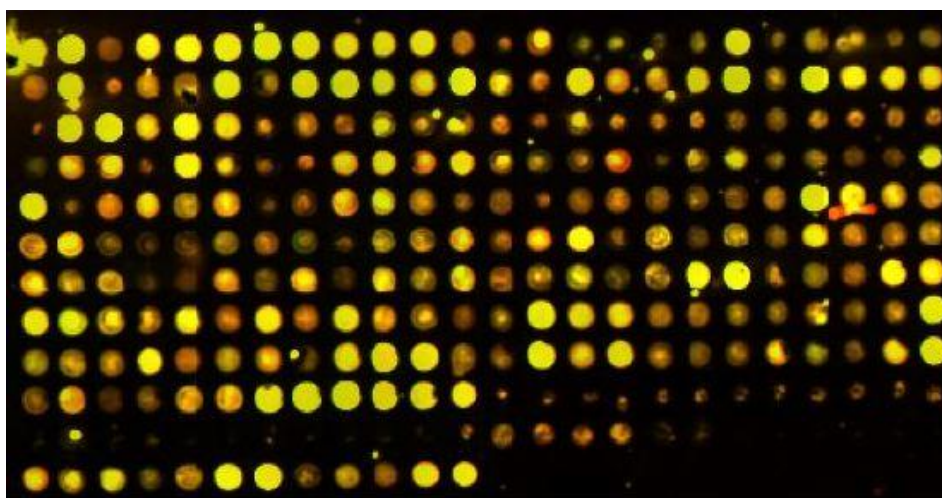
2.4.2. Εφαρμογή πλέγματος (gridding process)

Το πρώτο βήμα προεπεξεργασίας, είναι η εφαρμογή του πλέγματος στην εικόνα, ώστε ο ερευνητής να μπορεί εύκολα να χειρίζεται μια-μια τα στίγματα και να ανατρέχει σε οποιαδήποτε από αυτά επιθυμεί, ανά πάσα στιγμή. Με αυτό τον τρόπο γίνονται ευκολότερα τα επόμενα βήματα ανάλυσης [5].

2.4.3. Αξιολόγηση της Ποιότητας των Δεδομένων

Πριν τα δεδομένα των εικόνων υποστούν περαιτέρω ανάλυση, πρέπει να γίνουν κάποια βήματα αξιολόγησης των δεδομένων. Τα ανεπεξέργαστα δεδομένα πρέπει να είναι καλής ποιότητας, τα οποία να έχουν εκτιμηθεί, για την εξασφάλιση της ακεραιότητάς τους. Τα πρώτα ανεπεξέργαστα δεδομένα, θα πρέπει πάντα, να υπόκεινται σε κάποια μορφή τεχνικής διακύμανσης και επομένως, πρέπει να υποστούν κάποια προεπεξεργασία για να αφαιρεθούν όσες ανεπιθύμητες πηγές θορύβου υπάρχουν, προκειμένου να διασφαλιστεί ότι τα αποτελέσματα είναι στο υψηλότερο δυνατό επίπεδο ακρίβειας [5].

Σε επόμενο στάδιο αφαιρούνται όποια δεδομένα κρίθηκαν χαμηλής αξιοπιστίας και ακεραιότητας. Για παράδειγμα, στην εικόνα του Σχήματος 2.5. η δεύτερη και τρίτη από το τέλος σειρά στιγμάτων, περιέχουν στίγματα τα οποία είναι αμφισβητήσιμης αξιοπιστίας καθώς η ποιότητάς τους είναι πολύ χαμηλή.



Σχήμα 2.5.: Πραγματική εικόνα DNA

2.4.4. Κβαντοποίηση – Κανονικοποίηση

Ο σκοπός αυτού του βήματος είναι να προσαρμοστούν τα δεδομένα για την τεχνική τροποποίηση, λόγω των βιολογικών διαφορών μεταξύ των δειγμάτων. Πάντα θα υπάρχει μια ελαφριά ασυμφωνία μεταξύ των διαδικασιών υβριδισμού για κάθε εικόνα και οι διακυμάνσεις αυτές τείνουν να οδηγήσουν σε κλιμάκωση των διαφορών μεταξύ των συνολικών επιπέδων έντασης φθορισμού των διαφόρων

πινάκων. Για παράδειγμα, η ποσότητα του RNA σε ένα δείγμα, το χρονικό διάστημα για το οποίο ένα δείγμα υβριδοποιείται ή ο όγκος του δείγματος, μπορεί να προκαλέσουν σημαντικές διαφορές. Ακόμη και οι ανεπαίσθητες σωματικές διαφορές μεταξύ εικόνων ή μεταξύ των σαρωτών που χρησιμοποιούνται για να διαβάσουν οι εικόνες αυτές, μπορεί να έχουν επιπτώσεις [5].

2.4.5. Προσαρμογή των Αποτελεσμάτων σε Διεθνή Πρότυπα

Η υιοθέτηση διεθνών προτύπων θεωρείται ζωτικής σημασίας στο τομέα της επιστήμης, λόγω της σύγχυσης που δημιουργείται από τη χρήση διαφορετικών μονάδων. Ένα παρόμοιο θέμα αντιμετωπίζουν και οι ερευνητές που ασχολούνται με τις εικόνες DNA. Η ίδια αύξηση ή ελάττωση της γονιδιακής έκφρασης που παρατηρήθηκε σε δύο διαφορετικά εργαστήρια, θα μπορούσε πράγματι να είναι διαφορετική, ειδικά όταν χρησιμοποιούνται διαφορετικά πειραματικά πρωτόκολλα και μέθοδοι ανάλυσης δεδομένων. Χωρίς ένα πρότυπο, είναι σχεδόν αδύνατο να κριθεί η εγκυρότητα του αποτελέσματος μόνο με επιθεώρηση των αλλαγών υβριδοποίησης ή των ανεπεξέργαστων δεδομένων.

Λαμβάνοντας υπόψη αυτό το πρόβλημα, ο οργανισμός Microarray Gene Expression Data (MGED) Society, μια διεθνής πρωτοβουλία για την ανάπτυξη προτύπων για τις εικόνες DNA, πρότεινε το πρότυπο Minimum Information About a Microarray Experiment (MIAME) [8]. Οι ερευνητές έχουν αγκαλιάσει αυτό το πρότυπο και πολλές μεγάλες επιστημονικές κοινότητες απαιτούν τώρα την τήρηση του MIAME για κάθε νέα υποβολή. Επομένως, συνίσταται να επιβεβαιώνεται ότι, ο πειραματικός σχεδιασμός, τα πρωτόκολλα που ακολουθούνται αλλά και η ανάλυση των δεδομένων, συμμορφώνονται με το πρότυπο MIAME [12].

Τεχνικές Εφαρμογής Πλέγματος

3.1. Διαδικασία Εφαρμογής Πλέγματος – Αρχική Προσέγγιση

Η εφαρμογή πλέγματος, είναι ένα από τα σημαντικότερα στάδια επεξεργασίας για την ανάλυση εικόνων DNA. Το πλέγμα καταγράφει μια σειρά από άνισα διαστήματα, παράλληλων αλλά και κατακόρυφων, ως προς το περιεχόμενο της εικόνας, γραμμών, που αντιπροσωπεύουν ένα διδιάστατο (2D) πίνακα από στίγματα της εικόνας.

Γενικά οι μέθοδοι εφαρμογής πλέγματος μπορούν να διακριθούν σε τρία είδη. Οι χειροκίνητες-παραμετροποιήσιμες μεθοδολογίες με αναγκαία επίβλεψη από τον ερευνητή, ο οποίος χρειάζεται να εισάγει κάποιες παραμέτρους, οι ημιαυτόματες μέθοδοι και οι πλήρως αυτοματοποιημένες μέθοδοι, οδηγούμενες από τα ίδια τα δεδομένα (“data driven” methods) κατά τις οποίες η μοναδική είσοδος που απαιτείται είναι η ίδια η εικόνα και οι οποιεσδήποτε αναγκαίες παράμετροι, υπολογίζονται από τον ίδιο τον αλγόριθμο.

Τα προβλήματα σε αυτές τις μεθοδολογίες εφαρμογής πλέγματος, αρχίζουν όταν οι εικόνες DNA εμπεριέχουν θόρυβο στο υπόβαθρο (background) ή έχουν υποστεί παραμόρφωση. Οι ιδανικές εικόνες DNA (βλ. Σχήμα 2.3.), έχουν σταθερή γεωμετρία στιγμάτων, γνωστή ένταση φόντου με μηδενική αβεβαιότητα και άπειρη χωρική ανάλυση. Επίσης, τα στίγματα έχουν προκαθορισμένα σχήματα (μορφολογία) και

έχουν συνεχή ένταση φωτεινότητας η οποία: α) είναι διαφορετική από το υπόβαθρο (background), β) είναι ευθέως ανάλογη προς το βιολογικό φαινόμενο και γ) έχει μηδενική αβεβαιότητα. Για εικόνες πολλαπλών καναλιών, τα ίδια ισχύουν για το κάθε κανάλι, καθώς και ότι τα κανάλια είναι πλήρως ευθυγραμμισμένα.

Σε πραγματικά δεδομένα όμως, τέτοιες ιδανικές συνθήκες δυστυχώς αποτελούν ουτοπία, καθώς οι πραγματικές εικόνες DNA, υποφέρουν από θόρυβο και παραμορφώσεις. Υπό αυτές τις συνθήκες, κάποιιο πιο ευφυείς αλγόριθμοι, είναι απαραίτητοι για αντιμετώπιση τέτοιων προβλημάτων.

3.2. Υπάρχουσες Τεχνικές Εφαρμογής Πλέγματος

Σε αυτή την ενότητα, παρουσιάζονται επιγραμματικά χαρακτηριστικές τεχνικές εφαρμογής πλέγματος που χρησιμοποιούνται από τους ερευνητές και περιλαμβάνουν απλούς αλγόριθμους καθώς και κάποιες πιο σύνθετες και ευφυείς μεθοδολογίες.

3.2.1. Μέθοδος Προβολών

Η μέθοδος προβολών (projection pursuit method) είναι η πιο ευρέως διαδεδομένη μέθοδος εφαρμογής πλέγματος καθώς είναι και η απλούστερη. Σε τέτοιου είδους προβλήματα, η απλούστερη μέθοδος συνεπάγεται και μικρότερη πολυπλοκότητα, γεγονός το οποίο καθιστά, μεθοδολογίες που ικανοποιούν αυτές τις προδιαγραφές, να υπερτερούν σε σχέση με άλλες πολυπλοκότερες μεθόδους.

Λόγω του ότι τα πειράματα με εικόνες DNA γίνονται με βοήθεια ρομποτικών μηχανισμών, υπάρχει μεγάλη πιθανότητα η εικόνα μας να έχει υποστεί μια περιστροφή λόγω μη συμφωνίας των αξόνων του ρομπότ και της επιφάνειας εργασίας. Επομένως, σύμφωνα με την μέθοδο αυτή, υπολογίζεται πρώτα η γωνία κατά την οποία έχει περιστραφεί η εικόνα για να μπορεί να συνεχιστεί η μεθοδολογία. Ο υπολογισμός της γωνίας μπορεί να γίνει με εξαντλητικές μεθόδους οι οποίες αν και συνήθως αποφεύγονται, σε τέτοιες περιπτώσεις δεν είναι τόσο χρονοβόρες όσο αναμένεται θεωρητικά, αφού το εύρος των γωνιών που μπορεί να έχουμε είναι αρκετά μικρό, της τάξης έως $\pm 10^\circ$. Εναλλακτικές μέθοδοι υπολογισμού της γωνίας περιστροφής είναι μετασχηματισμοί όπως ο Radon ή Hough [11], που είναι η πιο γενική του μορφή, ή ακόμη με επαναληπτικές μεθόδους εφαρμοσμένες σε εικόνες αυτοσυσχέτισης [10].

Μετά την επαναφορά της εικόνας στο σωστό προσανατολισμό, εφαρμόζεται μια σειρά βημάτων προεπεξεργασίας για εξομάλυνση του θορύβου και διαγραφή των σφαλμάτων, με χρήση γκαουσιανών φίλτρων [11]

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (3.1)$$

ή/και φίλτρων ενδιάμεσης τιμής [11]

$$m = \text{median}\{X_{[m,n]}\} \quad (3.2)$$

Ο υπολογισμός των προβολών στον οριζόντιο και κατακόρυφο άξονα, γίνεται σύμφωνα με τις πιο κάτω σχέσεις:

$$I_{x\text{Projection}}(m,n) = \sum_{m=1}^M I(m,n) \quad (3.3.\alpha)$$

$$I_{y\text{Projection}}(m,n) = \sum_{n=1}^N I(m,n) \quad (3.3.\beta)$$

Σαρώνοντας στήλη προς στήλη όλη την εικόνα, αθροίζονται οι φωτεινότητες των εικονοστοιχείων (pixels) που συναντούνται στην εκάστοτε στήλη (Σχέση 3.3.α). Όταν η διαδικασία αυτή ολοκληρωθεί για κάθε στήλη της εικόνας, το αποτέλεσμα που παίρνει κανείς είναι ένας πίνακας γραμμή που αποτελεί την προβολή (projection) της εικόνας στον οριζόντιο άξονα και ονομάζεται οριζόντια προβολή. Αντίστοιχα υπολογίζεται και η κατακόρυφη προβολή της εικόνας.

Στο σημείο αυτό, είναι δυνατόν να εφαρμοστεί μια κατωφλιοποίηση στις προβολές, το οποίο γενικά είναι πολύ αισιόδοξο καθώς προϋποθέτει ακριβή υπολογισμό της γωνίας περιστροφής στο προηγούμενο βήμα. Αν έχουμε ακόμη και μια μικρή απόκλιση, πολύ πιθανό να πάρουμε λάθος αποτελέσματα. Μια πιο σωστή προσέγγιση είναι ο υπολογισμός των τοπικών μεγίστων στις δύο προβολές, ο οποίος θα μας δώσει τα σημεία από τα οποία θα σχεδιάσουμε τις γραμμές (οριζόντιες και κατακόρυφες) του πλέγματος. Ο αλγόριθμος αναζήτησης τοπικών μεγίστων, σαρώνει όλη την εικόνα και ψάχνει για κάθε προκαθορισμένων διαστάσεων περιοχή, το σημείο με την μέγιστη φωτεινότητα ενώ ταυτόχρονα εγγυάται ότι θα είναι το μοναδικό μέγιστο σε εκείνη την περιοχή.

Η χρήση των κατωφλίων (thresholds) αλλά και κάποιοι άλλοι παράμετροι όπως το διάστημα αναζήτησης για την γωνία περιστροφής, καθιστούν την μέθοδο των προβολών στην πρώτη κατηγορία μεθόδων, τις χειροκίνητες-παραμετροποιήσιμες μεθόδους.

Η μέθοδος αυτή είναι εξαιρετικά απλή, γεγονός που την καθιστά στις πρώτες θέσεις όσον αφορά την πολυπλοκότητα και τον χρόνο εκτέλεσης. Είναι αξιόπιστη για εικόνες που έχουν υποστεί μόνο περιστροφή ή γενικότερα παραμορφώσεις από αφφινικούς μετασχηματισμούς. Σε περιπτώσεις όμως, που η εικόνα μας έχει παραμορφωθεί από συνημιτονοειδή μετασχηματισμούς ή ακόμη αλλοιώσεις οπτικής (για παράδειγμα από φακούς), δυστυχώς δεν μπορεί να αντιμετωπίσει το πρόβλημα της εφαρμογής πλέγματος καθώς θα αποκλίνει από το βέλτιστο αποτέλεσμα.

3.2.2. Μέθοδος με Αναγνώριση “σημαντικών” στιγμάτων (guide spots)

Μια παραλλαγή της προηγούμενης μεθόδου, είναι η εφαρμογή του ιδίου του αλγορίθμου σε μια εικόνα, ενδιάμεσου σταδίου, η οποία περιλαμβάνει τα πιο “σημαντικές” στίγματα (guide spots) της αρχικής εικόνας DNA [9].

Η αναγνώριση των “σημαντικών” αυτών στιγμάτων, γίνεται με χρήση δύο ψηφιακών φίλτρων και αναζήτηση τοπικών μεγίστων. Το γραμμικό φίλτρο ταιριάσματος (linear matched filter) κατασκευάζεται σχηματίζοντας μια εικόνα προτύπου (S) με τα ιδανικά “σημαντικά” στίγματα. Το συγκεκριμένο φίλτρο (m) σαρώνει την εικόνα pixel προς pixel (m,n) της εικόνας S και η εικόνα εξόδου κανονικοποιείται με την μέση τιμή τοπικής φωτεινότητας ($\mu_{s[m,n]}$) [9].

$$\tilde{s}_{[m,n]} = s_{[m,n]} - \mu_{s[m,n]} \cdot I \quad (3.4)$$

Το τελικό αποτέλεσμα είναι το γινόμενο

$$R_{[m,n]}^M = \tilde{s}_{[m,n]} \cdot m \quad (3.5)$$

το οποίο αντιστοιχεί στην στατιστική συνδιακύμανση μεταξύ των εικόνων S και m. Σημεία υψηλών τιμών φωτεινότητας, δηλώνουν θέσεις πιθανών “σημαντικών” στιγμάτων (guide spots) [9].

Η κύρια ιδέα στην διάκριση των σημαντικών στιγμάτων ως προς τα υπόλοιπα στίγματα, είναι η ενίσχυση των “σημαντικών” στιγμάτων, θεωρώντας την απόκριση του γραμμικού φίλτρου ταιριάσματος σε μια γειτονιά 8 σημείων (8-connectivity). Σχηματίζεται το σύνολο $T_{[m,n]}$ το οποίο περιέχει τις αποκρίσεις $R_{[m,n]}$ και των γειτονιών των “σημαντικών” στιγμάτων [9].

$$T_{[m,n]} = \{R_{[m+k,n+l]}^M | k \in \{0, M_B, -M_B\} \cap l \in \{0, N_B, -N_B\}\} \quad (3.6)$$

Το αποτέλεσμα της διαδικασίας ενίσχυσης “σημαντικών” στιγμάτων (Guide Spot Location Amplification – GSLA) ($R_{[m,n]}^A$), υπολογίζεται ως το αποτέλεσμα του παρακάτω μη γραμμικού φίλτρου ενδιάμεσης τιμής [11]:

$$R_{[m,n]}^A = \text{median}\{T_{[m,n]}\} \quad (3.7)$$

3.2.3. Μέθοδος με χρήση Χαρτών απόστασης (distance maps)

Αφού προσανατολιστεί η εικόνα σύμφωνα με την γωνία πιθανής περιστροφής που μπορεί να είχε υποστεί, κατασκευάζονται δύο δυαδικές (binary) εικόνες. Η πρώτη περιέχει ένα δείκτη με γραμμή παράλληλη ως προς τον οριζόντιο (x) άξονα και αντίστοιχα η δεύτερη περιέχει ένα δείκτη για γραμμή παράλληλη ως προς τον κατακόρυφο (y) άξονα.

Στην συνέχεια, εφαρμόζεται ένας μετασχηματισμός απόστασης σε κάθε μια από τις δυαδικές εικόνες, χρησιμοποιώντας την μετρική του τετραγώνου (city block metric). Το αποτέλεσμα αυτής της διαδικασίας είναι μια εικόνα με διαβαθμίσεις του γκρι, όπου κάθε pixel δηλώνει την απόσταση μεταξύ κάθε pixel με ετικέτα 0 και του δείκτη. Η απόσταση των pixel υπολογίζεται σύμφωνα με την (3.8), όπου x_i , y_i είναι οι συντεταγμένες του κάθε pixel στην δυαδική εικόνα [10].

$$d_{min} = |x_1 - y_1| + |x_2 - y_2| \quad (3.8)$$

Μετά την εφαρμογή του μετασχηματισμού απόστασης, ακολουθεί μια διαδικασία κανονικοποίησης για να διακρίνονται μεταξύ τους οι γραμμές σε διακριτούς ακέραιους αριθμούς. Το αποτέλεσμα του μετασχηματισμού απόστασης είναι μια εικόνα με διαβαθμίσεις του γκρι, η οποία ονομάζεται χάρτης αποστάσεων (distance map). Οι χάρτες απόστασης είναι ένα σύνολο παράλληλων γραμμών που έχουν τον ίδιο ακέραιο αριθμό.

3.2.4. Προσέγγιση “Ορειβασίας σε Λόφο” (Hill-Climbing Approach)

Η συγκεκριμένη μέθοδος, όντας πλήρως αυτοματοποιημένη, εν αντιθέσει με άλλες μεθόδους, δεν κάνει υποθέσεις για τον αριθμό των σειρών και στηλών από στίγματα, το μέγεθος τους ούτε για τα όρια του πλέγματος.

Σύμφωνα με τη μέθοδο αυτή, το πρόβλημα χωρίζεται σε δύο μέρη. Αρχικά, υπολογίζεται το μέγεθος (ακτίνα) των στιγμάτων ξεκινώντας από r τυχαία pixels στην εικόνα. Σε επόμενο στάδιο, αναγνωρίζονται τα στίγματα στην εικόνα χρησιμοποιώντας την πληροφορία του μεγέθους του στίγματος, από το προηγούμενο βήμα. Με μια αναδρομική διαδικασία διορθώνεται η αρχική εκτίμηση του μεγέθους των στιγμάτων, με εφαρμογή ενός αλγορίθμου αναζήτησης τοπικών μεγίστων (βλ. παράγραφο 3.2.1. για τη λειτουργία του συγκεκριμένου αλγόριθμου), σε μια περιοχή ανάλογα του μεγέθους των στιγμάτων που υπολογίστηκε προηγουμένως [7].

Αναλυτικότερα, η μέθοδος αυτή παίρνει ως αρχική είσοδο ένα μέρος της εικόνας DNA και υπολογίζει την “ενέργεια” της, πολλαπλασιάζοντας το εκάστοτε pixel με ένα βάρος υπολογιζόμενο από μια παραμετροποιημένη κατανομή πιθανοτήτων, της οποίας η τυχαία μεταβλητή είναι η απόσταση από το κέντρο της εικόνας μέχρι το αντίστοιχο pixel. Στην συγκεκριμένη μέθοδο, προτείνονται τρεις διαφορετικές κατανομές: η ομοιόμορφη, η εκθετική και η κανονική [7].

Στην συνέχεια, ένας αλγόριθμος σαρώνει την εικόνα προς αναζήτηση κορυφών (peaks) κατά μήκος μιας οριζόντιας γραμμής, αρχικά προς τα αριστερά και μετά προς τα δεξιά, αποθηκεύοντας τα αποτελέσματα για τις οριζόντιες γραμμές σε ένα διάνυσμα h_{peak} . Ένας άλλος παρόμοιος αλγόριθμος αναζητά κορυφές σε μια κατακόρυφη γραμμή, αρχικά προς τα πάνω και μετά προς τα κάτω και αποθηκεύει τα αντίστοιχα αποτελέσματα σε ένα αντίστοιχο διάνυσμα v_{peak} [7].

Αφού βρεθούν τα στίγματα με την μέγιστη “ενέργεια”, τόσο στις οριζόντιες όσο και στις κατακόρυφες γραμμές για τα πρώτα r τυχαία σημεία, το ίδιο γίνεται για όλη την εικόνα. Για να μπορεί αυτό να γίνει εφικτό, υπολογίζεται το πλάτος των στιγμάτων, αφαιρώντας τις διαδοχικές γραμμές κορυφών που είναι αποθηκευμένα στα διανύσματα h_{peak} και v_{peak} [7].

$$\hat{\mu}_i = \frac{1}{|h_{peak}|} \sum_{j=1}^{|h_{peak}|-1} h_{peak}[j] - h_{peak}[j+1] \quad (3.9)$$

Το τελικό αποτέλεσμα είναι η μέση τιμή των μέσων τιμών των γραμμών.

Κατασκευάζοντας ένα ιστόγραμμα για τις οριζόντιες γραμμές και ένα για τις κατακόρυφες, υπολογίζονται οι θέσεις των οριζόντιων και κατακόρυφων γραμμών του πλέγματος, με την βοήθεια των αντίστοιχων ιστογραμμάτων. Με τις πληροφορίες αυτές το πλέγμα μπορεί πλέον να σχεδιαστεί [7].

Αν και πλήρως αυτοματοποιημένη, χωρίς καθόλου υποθέσεις για σημαντικές παραμέτρους του συστήματος, η μέθοδος αυτή αντιμετωπίζει προβλήματα σε εικόνες που έχουν υποστεί παραμόρφωση από μετασχηματισμούς πέραν του αφφινικού.

3.2.5. Μπεϊσιανό ταίριασμα πλέγματος (Bayesian Grid Matching)

Μια πιο ευφυής μέθοδος για εφαρμογή πλέγματος σε εικόνες DNA, είναι η αυτοματοποιημένη μέθοδος που χρησιμοποιεί την Μπεϊσιανή θεωρία (Bayesian Theory) για τη βέλτιστη προσαρμογή του πλέγματος στην εικόνα.

Γενικά, η μπεϊσιανή θεωρία κάνει χρήση των εκ των προτέρων γνώσεων (prior) και με βοήθεια μετρούμενων μεγεθών, διορθώνει την αρχική εκτίμηση, εξάγοντας μια εκ των υστέρων εκτίμηση (posterior) [4].

$$(\text{εκ των υστέρων}) = \frac{(\text{πιθανοφάνεια}) * (\text{εκ των προτέρων})}{(\text{τεκμήρια})} \quad (3.10)$$

Αρχικά, δηλώνονται κάποιες έννοιες όπως: το σύνολο των κόμβων ενός γράφου (S), το σύνολο των τόξων – ακμών (L), τη γειτονιά ενός κόμβου (N) (γειτονιά ενός κόμβου είναι οι κόμβοι με τους οποίους ενώνεται – το l_{ij} δηλώνει ότι οι κόμβοι i και j ενώνονται) και το σύνολο των τόξων που δηλώνουν την παραμόρφωση του γράφου (T) ώστε T_{ij} είναι το διάστημα μεταξύ των κόμβων i και j [6].

Σύμφωνα με το θεώρημα των Hammersley-Clifford, με χρήση τυχαίων πεδίων Gibbs (Gibbs Random Fields), τα οποία στην γενική τους μορφή είναι Μαρκοβιανά τυχαία πεδία (Markov Random Fields), υπολογίζονται οι υπό συνθήκη πυκνότητες $P(g_i | \{g_j, s_j \neq s_i\})$, οι οποίες ακολουθούν κανονική κατανομή $N(u_i, S_i)$. Με αυτά, δηλώνεται η εκ των προτέρων γνώση για το γράφο των κόμβων [6].

Για την εκ των προτέρων γνώση του πλέγματος, μοντελοποιείται ένα δισδιάστατο τετράγωνο πλέγμα. Στην περίπτωση των δύο διαστάσεων, οι θέσεις των κόμβων του πλέγματος γράφονται ως $g_i = (g_i^{(x)}, g_i^{(y)})$.

Οι δυνάμεις “συγκράτησης” μεταξύ των κόμβων του πλέγματος μοντελοποιούνται σε μορφή “δυνάμεων ελατηρίου” με την πιο κάτω εξίσωση.

$$\Sigma_{ij} = (T_{ij}, \hat{T}_{ij}) \begin{pmatrix} \beta_{ij, long}^{-1} & 0 \\ 0 & \beta_{ij, trans}^{-1} \end{pmatrix} \begin{pmatrix} T_{ij} \\ \hat{T}_{ij} \end{pmatrix} \quad (3.11)$$

Στην συνέχεια, συνδυάζονται οι εκ των προτέρων γνώσεις για τους κόμβους του γράφου και του πλέγματος, ώστε να συντελέσουν μια ενιαία.

Μια εκτίμηση των υπολοίπων παραμέτρων γίνεται με χρήση ενός εκτιμητή ψευδοπιθανοφάνειας, η οποία λόγω του γκαουσιανού υπόβαθρου, υποβιβάζεται σε μέθοδο ελαχίστων τετραγώνων.

Η εικόνα παρατήρησης αντιπροσωπεύεται από τον όρο Y . Το μοντέλο παρατήρησης $P(Y|G)$ περιγράφει την πιθανότητα να παρατηρηθεί ο όρος Y , δεδομένου του όρου G .

Η εκ των υστέρων κατανομή του μοντέλου του πλέγματος είναι επίσης τύπου Μαρκοβιανών τυχαίων πεδίων (MRF) και κατοπτρίζει μια αναλογία μεταξύ της κανονικότητας του πλέγματος και της αξιοπιστίας της εικόνας μας. Η εκ των υστέρων κατανομή γράφει ως εξής.

$$P(G, D|Y) = P(G|D) \cdot P(Y|G, D) \quad (3.12)$$

Στο συγκεκριμένο μοντέλο, οι παράμετροι του πλέγματος μπορούν να ελεγχθούν. Η αξιοπιστία των δεδομένων ελέγχεται από τον όρο $P(Y|G)$ και το ποσό της κανονικότητας στο πλέγμα ελέγχεται από τον όρο $P(G|D)$. Η εκ των υστέρων κατανομή μεγιστοποιείται δειγματολειπώντας την εκ των υστέρων κατανομή, χρησιμοποιώντας το Metropolis Sampling Scheme [6].

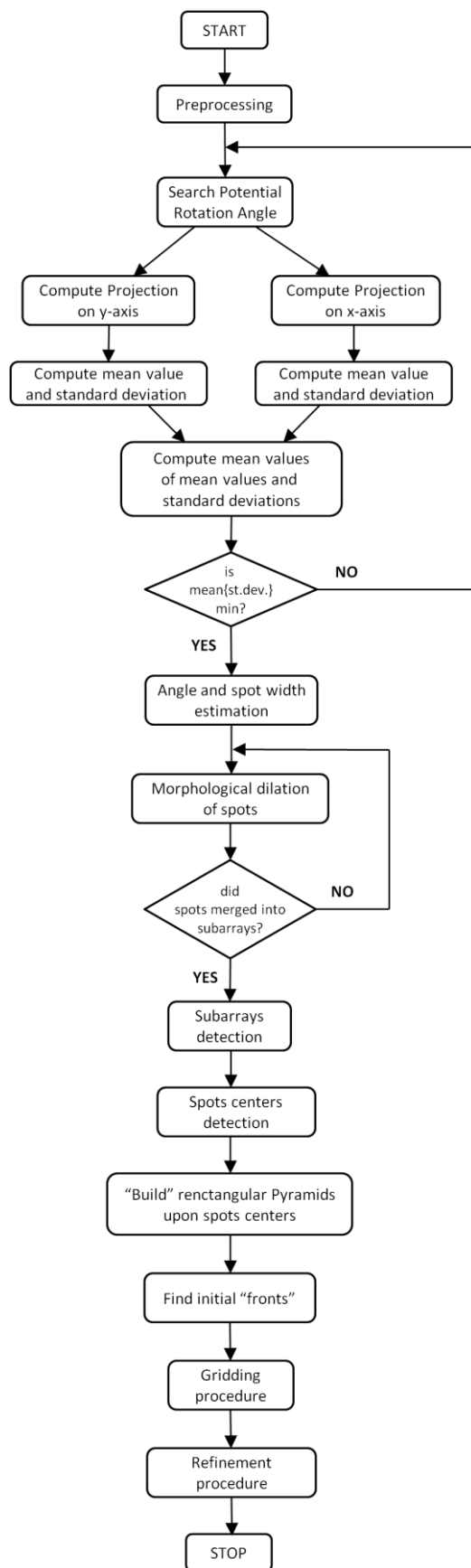
Για μεγιστοποίηση της (3.10) εφαρμόζεται ένα Simulated Annealing (SA) scheme, το οποίο είναι μια Μαρκοβιανή Αλυσίδα Monte Carlo (Markov Chain Monte Carlo – MCMC) η οποία είναι μια μέθοδος βελτιστοποίησης που ενεργεί με τροποποιήσιμες θερμοκρασίες, όπως για παράδειγμα μια ανομοιογενής μαρκοβιανή αλυσίδα. Η μεγιστοποίηση της (3.12) δίνει το επιθυμητό βέλτιστο αποτέλεσμα [6].

Μεθοδολογία Αυτόματης Εφαρμογής Πλέγματος

4.1. Η Μεθοδολογία Αυτόματης Εφαρμογής Πλέγματος

Μια εξήγηση του προβλήματος εφαρμογής πλέγματος σε εικόνες DNA, είναι η χάραξη καμπύλων στα διαστήματα, ενδιάμεσα των στιγμάτων. Αυτό είναι προφανές ότι θα απομονώσει τα στίγματα μεταξύ τους, το οποίο είναι και το τελικό επιθυμητό αποτέλεσμα. Στη προσπάθεια για ανάπτυξη μιας αυτόματης μεθοδολογίας για εφαρμογή πλέγματος σε εικόνες DNA, η αρχική ιδέα εμπνεύστηκε από την φύση. Το φυσικό σύστημα, από το οποίο προήλθε η αρχική έμπνευση για την υλοποίηση της προτεινόμενης μεθοδολογίας, είναι οι λόφοι και τα ρυάκια που ρέουν μεταξύ τους. Το νερό πάντοτε αναζητά σημεία με το μικρότερο δυναμικό για να κινηθεί προς αυτά. Αυτό ακριβώς είναι η κύρια ιδέα πίσω από την αυτόματη μεθοδολογία που υλοποιήθηκε στα πλαίσια αυτής της διπλωματικής εργασίας.

Αρχικά, το θέμα προσεγγίστηκε προσπαθώντας να οδηγηθούν οι καμπύλες του πλέγματος ενδιάμεσα των στιγμάτων όπως ακριβώς είναι στην εικόνα. Αν και προηγουμένως εφαρμόστηκαν διάφορες τεχνικές για απαλοιφή του θορύβου και φιλτραρίσματα εξομάλυνσης, αυτή η προσέγγιση απέτυχε. Προκειμένου όμως οι λόφοι να προσεγγιστούν καλύτερα, στη θέση των στιγμάτων προτιμήθηκε η εμφάνιση τετραγωνικών πυραμίδων, με τις οποίες θα μπορούσε να εκμεταλλευτεί κανείς το γεγονός ότι έχουν παράλληλες τις πλευρές τους σε κάθε επίπεδο.



Σχήμα 4.1.: Μπλοκ διάγραμμα της προτεινόμενης μεθοδολογίας

Οι τετραγωνικές αυτές πυραμίδες, έπρεπε να σχεδιαστούν ακριβώς στην θέση που αρχικά βρίσκονταν τα στίγματα. Για να επιτευχθεί κάτι τέτοιο έπρεπε να αναγνωριστούν με σαφήνεια και ακρίβεια τα κέντρα των στιγμάτων, ούτως ώστε να είναι εφικτή η σχεδίαση των πυραμίδων στις σωστές συντεταγμένες.

Σε αυτό το σημείο, πρέπει να δοθεί μια πιο λεπτομερής περιγραφή των βημάτων που απαρτίζουν την προτεινόμενη μεθοδολογία, ώστε ο αναγνώστης να είναι σε θέση να κατανοήσει πλήρως την όλη διαδικασία. Στο Σχήμα 4.1, παρουσιάζεται το μπλοκ διάγραμμα της προτεινόμενης μεθοδολογίας αυτόματης εφαρμογής πλέγματος.

4.1.1. Αρχικά στάδια προεπεξεργασίας της εικόνας DNA

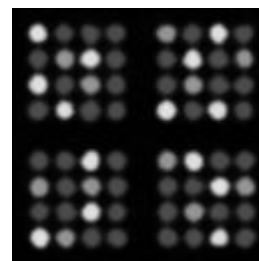
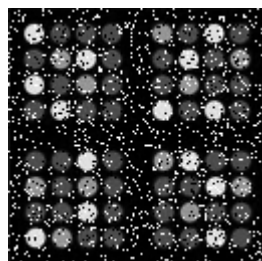
Σαν πρώτο βήμα στην επεξεργασία εικόνων, συνηθίζεται να εφαρμόζεται ένα φίλτρο εξομάλυνσης της εικόνας, όπως το γκαουσιανό φίλτρο (Gaussian filter). Ένα πέρασμα της εικόνας από γκαουσιανό φίλτρο (Σχέση 4.1.) τυπικής απόκλισης $s = 0.9$, δίνει πολύ ικανοποιητικά αποτελέσματα. Με αυτό το τρόπο, η εικόνα ετοιμάζεται για απαλοιφή του θορύβου που καταλαμβάνει το υπόβαθρο (background).

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (4.1)$$

Στο σημείο αυτό, εφαρμόζεται ένα φίλτρο ενδιάμεσης τιμής (median filter) το οποίο αποδεικνύεται το βέλτιστο για αντιμετώπιση θορύβου τύπου “salt and pepper”. Ο θόρυβος αυτός εμφανίζει εικονοστοιχεία (pixels) υψηλής φωτεινότητας (“salt”) και χαμηλής (“pepper”) σε διάσπαρτα σημεία της εικόνας. Το φίλτρο ενδιάμεσης τιμής (Σχέση 4.2) καθώς σαρώνει ολόκληρη την εικόνα, αναζητά σε μια προκαθορισμένων διαστάσεων, γειτονιά εικονοστοιχείων (pixels) το σημείο με την ενδιάμεση τιμή και αντικαθιστά με αυτό, το κεντρικό εικονοστοιχείο στο οποίο εφαρμόζεται.

$$m = \text{median}\{X_{[m,n]}\} \quad (4.2)$$

Στο Σχήμα 4.2.α. παρουσιάζεται μια περίπτωση εικόνας με τυχαίο θόρυβο τύπου “salt and pepper” πυκνότητας 15%, ενώ παράλληλα στο Σχήμα 4.2.β. παρουσιάζεται το αποτέλεσμα μετά την αποθορυβοποίησης της, με χρήση ενός γκαουσιανού φιλτραρίσματος και φίλτρου ενδιάμεσης τιμής.



Σχήμα 4.2.α.: Εικόνα DNA με θόρυβο Σχήμα 4.2.β.: Αποτέλεσμα αποθορυβοποίησης

4.1.2. Εκτίμηση της γωνίας περιστροφής και της ακτίνας των στιγμάτων

Η εκτίμηση της γωνίας περιστροφής της εικόνας DNA, είναι ένα πολύ σημαντικό στάδιο, πάνω στο οποίο βασίζονται τα επόμενα βήματα επεξεργασίας. Αν δεν υπολογιστεί με ικανοποιητική ακρίβεια η γωνία περιστροφής, διακυβεύεται η επιτυχία της όλης μεθοδολογίας.

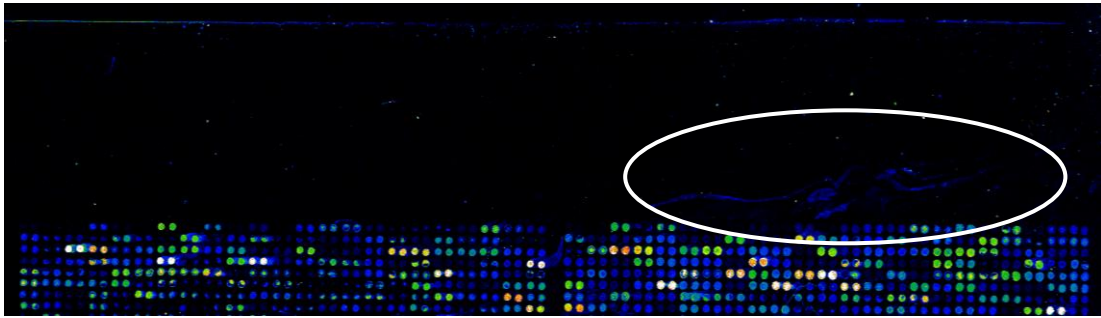
Όπως αναφέρθηκε σε προηγούμενο σημείο, η εκτίμηση της γωνίας περιστροφής μπορεί να γίνει με χρήση κάποιων μετασχηματισμών όπως Radon/Hough [11] ή με εξαντλητικές μεθόδους, αναζητώντας την σωστή γωνία σε ένα προκαθορισμένο διάστημα τιμών. Λόγω του τρόπου δημιουργίας των εικόνων DNA, το διάστημα $[-10^\circ, 10^\circ]$ είναι ικανοποιητικό. Πολύ σημαντικό είναι και το βήμα αύξησης της γωνίας αναζήτησης, καθώς όσο μεγαλύτερο είναι, τόσο γρηγορότερη γίνεται η αναζήτηση κοστίζοντας όμως στο θέμα της ακρίβειας. Αντίστοιχα, όσο μικρότερο είναι το βήμα αναζήτησης, τόσο μεγαλύτερη ακρίβεια επιτυγχάνεται με αντίκτυπο στην ταχύτητα της μεθόδου. Το γεγονός όμως ότι ο αλγόριθμος αναζήτησης είναι αρκετά απλός και το διάστημα αναζήτησης σχετικά μικρό, ένα βήμα αναζήτησης της γωνίας της τάξης του ενός δεκάτου (0.1) είναι μια πολύ καλή επιλογή.

Η διαδικασία εκτίμησης της γωνίας περιστροφής έχει ως εξής. Επιλέγεται όλη η εικόνα ή ένα τυχαίο μέρος της εικόνας DNA, το οποίο περιστρέφεται στο διάστημα αναζήτησης και σε κάθε βήμα υπολογίζονται οι προβολές σύμφωνα με τις Σχέσεις 4.3.α. και 4.3.β.

$$I_{xProjection}(m, n) = \sum_{m=1}^M I(m, n) \quad (4.3.α)$$

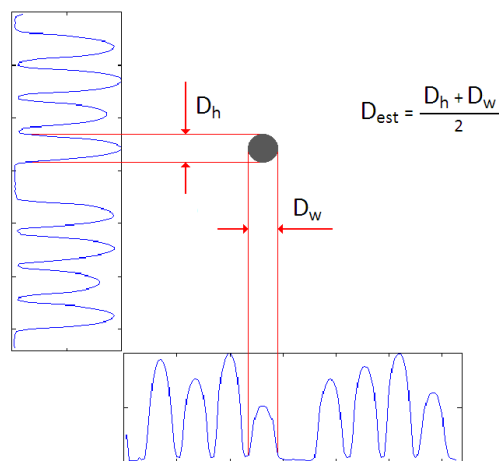
$$I_{yProjection}(m, n) = \sum_{n=1}^N I(m, n) \quad (4.3.β)$$

Αφού αρχικά οι προβολές κανονικοποιηθούν στο διάστημα $[0,1]$, καταωφλιοποιούνται στο 0.2 της μέγιστης τιμής φωτεινότητας της εικόνας. Η επιλογή του συγκεκριμένου καταωφλίου είναι αποδεκτή αν δεχτεί κανείς ότι ο θόρυβος στην εικόνα έχει φωτεινότητα μικρότερη από το ένα πέμπτο ($1/5$) της μέγιστης τιμής της φωτεινότητας στην εικόνα. Ο θόρυβος στην περίπτωση αυτή δεν περιλαμβάνει περιπτώσεις όπως του τύπου “salt and pepper”, αφού τέτοιου είδους θόρυβοι αντιμετωπίζονται στο πρώτο στάδιο προεπεξεργασίας της εικόνας DNA. Οι θόρυβοι που κλίνεται κανείς να αντιμετωπίσει σε αυτό το στάδιο, είναι από σφάλματα που προέκυψαν κατά την διάρκεια του πειράματος. Στο Σχήμα 4.3. παρουσιάζεται ένα κομμάτι εικόνας DNA που περιέχει θόρυβο λόγω εσφαλμένης υβριδοποίησης της επιφάνειας της κατά την διάρκεια του πειράματος. Αυτού του είδους θόρυβοι, προκαλούν σοβαρά προβλήματα, ειδικά στον υπολογισμό των προβολών της εικόνας αφού συνήθως έχουν υψηλές τιμές φωτεινότητας.



Σχήμα 4.3.: Εικόνα DNA με θόρυβο λόγω εσφαλμένης υβριδοποίησης

Ένα πολύ σημαντικό σημείο, είναι και η εκτίμηση της ακτίνας των στιγμάτων που περιέχει η εικόνα DNA. Στις πλείστες μεθόδους, γίνονται αρχικές υποθέσεις για την ακτίνα των στιγμάτων, στις οποίες βασίζονται τα επόμενα βήματα επεξεργασίας. Για μια αυτόματη και πιο αξιόπιστη εκτίμηση της ακτίνας, η προτεινόμενη μέθοδος υπολογίζει την ακτίνα των στιγμάτων από δεδομένα της εικόνας, χωρίς καμία απολύτως υπόθεση. Επομένως, για την εκτίμηση της ακτίνας των στιγμάτων, στην κάθε μια κατωφλιοποιημένη προβολή υπολογίζεται και το πλάτος του κάθε διαδοχικού παλμού τόσο στην οριζόντια όσο και στην κατακόρυφη προβολή. Στην συνέχεια, υπολογίζεται η μέση τιμή για κάθε προβολή, ενώ η μέση τιμή των δύο μέσων αυτών τιμών αντιστοιχεί στην εκτίμηση της διαμέτρου των στιγμάτων για την συγκεκριμένη περιστρεφόμενη εικόνα. Στο Σχήμα 4.4, παρουσιάζεται σχηματικά η εκτίμηση της ακτίνας ενός από τα στίγματα της εικόνας.

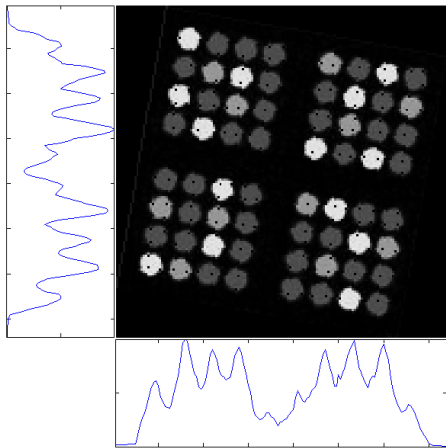


Σχήμα 4.4.: Εκτίμηση της διαμέτρου ενός στίγματος

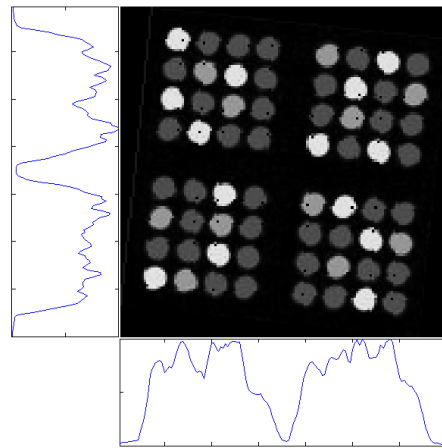
Το κριτήριο επιλογής της σωστής γωνίας περιστροφής, η οποία εμμέσως δίνει και την σωστή εκτίμηση της ακτίνας των στιγμάτων, είναι η κανονικότητα και η ασυνέχεια που εμφανίζεται στις κατωφλιοποιημένες προβολές. Με μαθηματικές παραστάσεις, αυτό μπορεί να εκφραστεί ως η τυπική απόκλιση (Σχέση 4.4), της κανονικής κατανομής, η οποία κατασκευάζεται με δεδομένα τις μέσες τιμές των παλμών στην εκάστοτε προβολή.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}, \quad \text{όπου} \quad \mu = \frac{1}{N} \sum_{i=1}^N x_i^2 \quad (4.4)$$

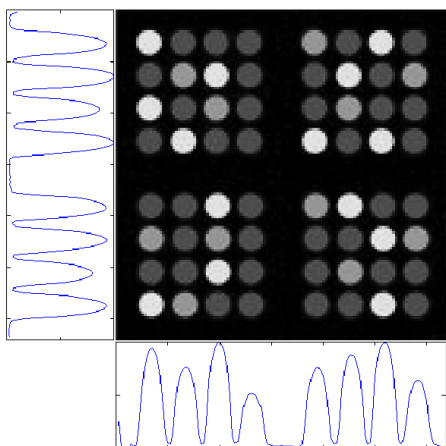
Η γωνία που δίνει την μικρότερη τυπική απόκλιση σύμφωνα με την Σχέση 4.4, είναι η σωστότερη επιλογή. Αυτό φαίνεται πολύ καλά στο πιο κάτω σχήμα όπου παρουσιάζονται διαφορετικά βήματα του αλγορίθμου αναζήτησης της γωνίας περιστροφής μαζί με τις προβολές και την τυπική απόκλιση των πλατών τους. Είναι εμφανές ότι η περίπτωση με την μικρότερη τυπική απόκλιση (Σχήμα 4.5.γ. με τυπική απόκλιση: $\sigma=1.8771$) δίνει την σωστή εκτίμηση της γωνίας περιστροφής της εικόνας DNA. Για ευκολία υλοποίησης και παρουσίασης των πιο κάτω αποτελεσμάτων, επειδή η εικόνα εισόδου ήταν μια ιδανική εικόνα DNA με μηδενική γωνία περιστροφής, ο αλγόριθμος αναζητούσε με ένα μεγάλο βήμα 2° σε διάστημα $[-10, +10]$ για μεγαλύτερη ταχύτητα εκτέλεσης.



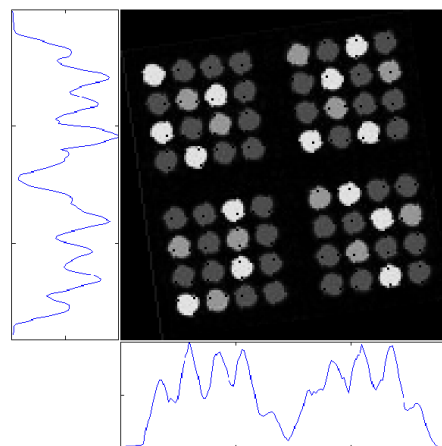
Σχήμα 4.5.α: Βήμα $N = 1$ αναζήτησης γωνίας περιστροφής ($\sigma = 44.6499$)



Σχήμα 4.5.β: Βήμα $N = 3$ αναζήτησης γωνίας περιστροφής ($\sigma = 41.0172$)



Σχήμα 4.5.γ: Βήμα $N = 5$ αναζήτησης γωνίας περιστροφής ($\sigma = 1.8771$)



Σχήμα 4.5.δ: Βήμα $N = 10$ αναζήτησης γωνίας περιστροφής ($\sigma = 19.8742$)

4.1.3. Αναζήτηση υποεικόνων (subarrays) και εύρεση των ορίων τους

Λόγω της φύσης των εικόνων DNA και των βημάτων που ακολουθήθηκαν κατά την διάρκεια των πειραμάτων, οι εικόνες DNA, στις πλείστες περιπτώσεις αποτελούνται από πλήθος υποεικόνων μέσα στην ίδια την εικόνα. Πολλές φορές η εικόνα τυπώνεται πολλαπλές φορές, οπότε στην ίδια επιφάνεια μπορεί να υπάρχουν περισσότερες από μια εικόνες DNA. Αυτό αναδεικνύει την ανάγκη για ένα αρχικό διαχωρισμό αυτών των υποεικόνων ώστε να επεξεργαστούν ως αυτόνομες εικόνες.

Ο τρόπος αναζήτησης των υποεικόνων που προτείνεται στα πλαίσια της διπλωματικής εργασίας, είναι η επαναλαμβανόμενη μορφολογική διαστολή των στιγμάτων, μέχρι την ένωση σε ένα μεγαλύτερο όγκο πληροφορίας ο οποίος στην ουσία είναι η υποεικόνα. Η Σχέση 4.5.α αφορά δυαδικές (binary images) εικόνες, ενώ η Σχέση 4.5.β αφορά εικόνες με φωτεινότητες διαβαθμίσεων του γκρι (grayscale images) [1].

$$A \oplus B = \bigcup_{b \in B} A_b = B \oplus A = \bigcup_{a \in A} B_a \quad (4.5.α)$$

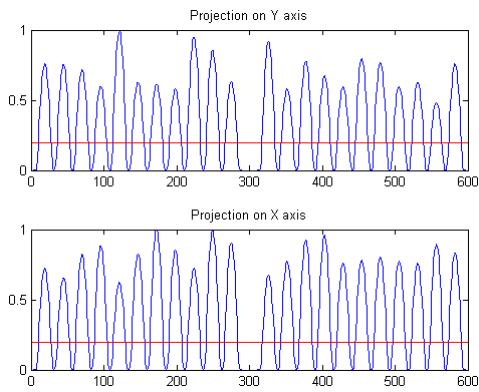
$$(f \oplus b)(x) = \sup_{y \in E} [f(y) + b(x - y)] = \sup_{z \in E} [f(x - z)] \quad (4.5.β)$$

Αναλυτικότερα, εφαρμόζεται μια σειρά επαναλαμβανόμενων μορφολογικών διαστολών στην εικόνα με ένα μορφολογικό τελεστή δίσκο ακτίνας 2 εικονοστοιχείων (pixels) και σε κάθε βήμα υπολογίζονται οι προβολές της διεσταλμένης εικόνας. Σε κάθε επανάληψη αυτής της διαδικασίας, γίνεται κατωφλιοποίηση στο ένα πέμπτο (1/5) της μέγιστης φωτεινότητας της εικόνας και υπολογίζεται το πλήθος των υποεικόνων που βρέθηκαν, σύμφωνα με τη Σχέση 4.6, η οποία χρησιμοποιεί την πληροφορία των θέσεων των κατακόρυφων και οριζόντιων γραμμών του πλέγματος για τις υποεικόνες.

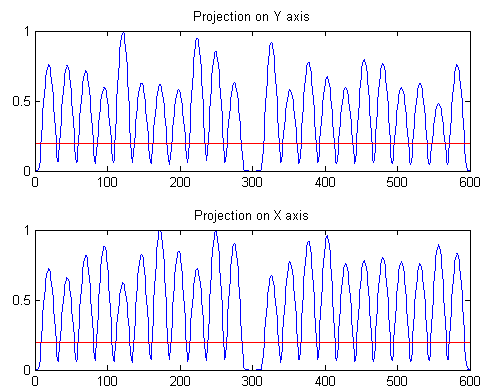
$$NumOfSubArrays = (N_{οριζοντιων γρ.} - 1) \cdot (N_{κατακορυφων γρ.} - 1) \quad (4.6)$$

Το κατώφλι (0.2) σε αυτή τη περίπτωση, δεν παίζει κανένα ρόλο αφού με τις συνεχόμενες μορφολογικές διαστολές, όποια τιμή και αν είχε το κατώφλι, τα στίγματα σε κάποιο βήμα της επανάληψης θα ενώνονταν και η κατωφλιοποίηση θα χώριζε τις υποεικόνες. Ο λόγος που επιλέχθηκε το ένα πέμπτο (1/5) για κατώφλι, είναι η υπόθεση που έγινε και σε προηγούμενο βήμα, ότι ο θόρυβος στο υπόβαθρο της εικόνας DNA (από εσφαλμένη υβριδοποίηση) έχει φωτεινότητα μικρότερη του ενός πέμπτου (1/5) της μέγιστης φωτεινότητας της εικόνας.

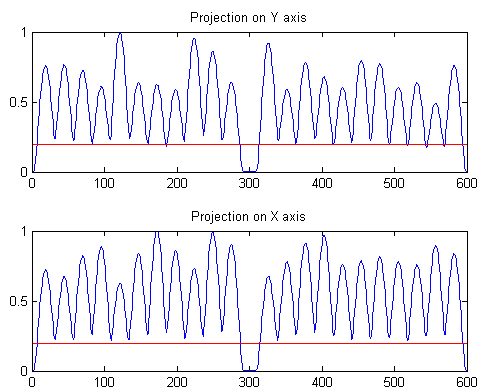
Στα Σχήματα 4.6.α ως και 4.6.δ, παρουσιάζονται οι κατακόρυφες (πάνω) και οι οριζόντιες (κάτω) προβολές της εικόνας για τα πρώτα τέσσερα αντίστοιχα βήματα του αλγορίθμου αναζήτησης των υποεικόνων, αφού τόσα χρειάζεται ο αλγόριθμος για να ανιχνεύσει τις υποεικόνες στην εικόνα DNA.



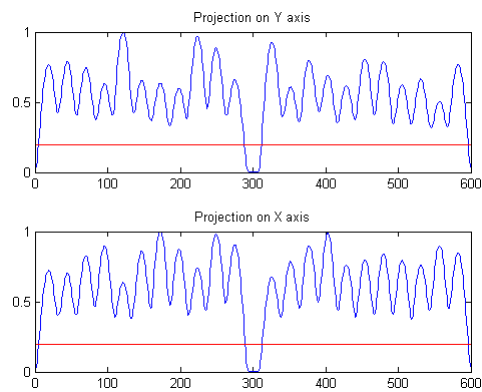
Σχήμα 4.6.α.: Βήμα N=1 αναζήτησης υποεικόνων



Σχήμα 4.6.β.: Βήμα N=2 αναζήτησης υποεικόνων

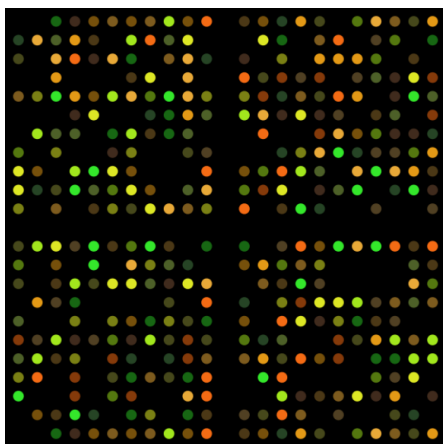


Σχήμα 4.6.γ.: Βήμα N=3 αναζήτησης υποεικόνων

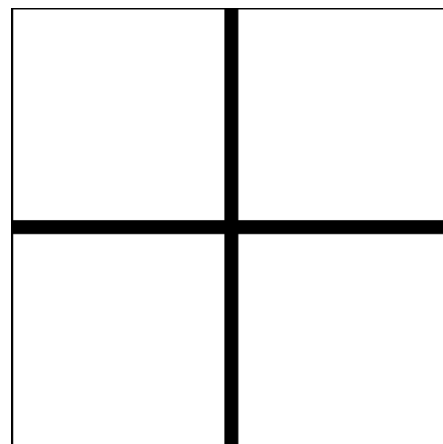


Σχήμα 4.6.δ.: Βήμα N=4 αναζήτησης υποεικόνων

Στο Σχήμα 4.7.β. παρουσιάζεται το αποτέλεσμα του αλγορίθμου ανίχνευσης υποεικόνων, με εικόνα εισόδου μια ιδανική εικόνα DNA (Σχήμα 4.7.α), η οποία παρουσιάζει σε κάποια σημεία κενά στίγματα, για απόδειξη λειτουργίας του αλγορίθμου σε πραγματικές εικόνες DNA με αντίστοιχες συνθήκες.



Σχήμα 4.7.α.: Εικόνα εισόδου για ανίχνευση υποεικόνων



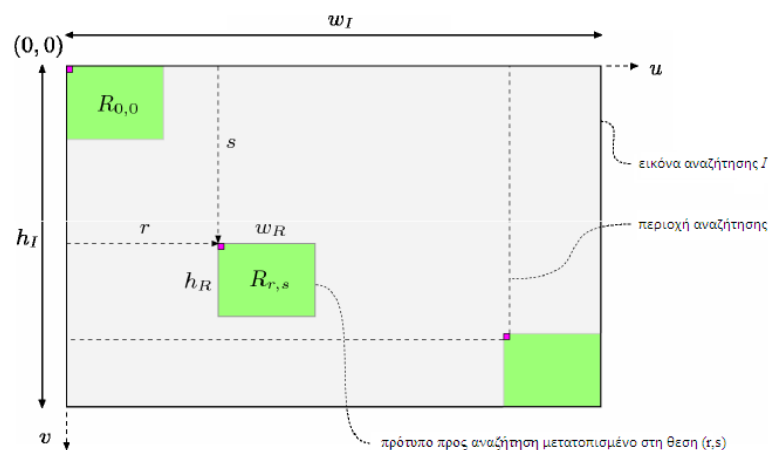
Σχήμα 4.7.β.: Επιτυχής ανίχνευση των υποεικόνων

Έχοντας τις προβολές σε μορφή τετραγωνικών παλμών, ανιχνεύονται οι θέσεις όπου υπάρχει εναλλαγή από 0 σε 1 και από 1 σε 0. Οι θέσεις αυτές αντιστοιχούν στις ακμές των λευκών τετραγώνων του Σχήματος 4.7.β. Με γνωστές τις εναλλαγές αυτές, υπολογίζονται οι δείκτες των γραμμών που διαχωρίζουν τις υποεικόνες, βρίσκοντας τα σημεία ανάμεσα των σημείων των εναλλαγών. Για τις ακραίες γραμμές διαχωρισμού των υποεικόνων, υπολογίζονται τα σημεία ανάμεσα στην εκάστοτε ακραία γραμμή και την αντίστοιχη ακμή της εικόνας.

4.1.4. Αναγνώριση των στιγμάτων και εύρεση των κέντρων τους

Γενικά, το πρόβλημα της αναζήτησης συγκεκριμένων προτύπων σε ένα ευρύτερο σύνολο που τις περιέχει, στην προκειμένη περίπτωση μια εικόνα DNA, είναι πολύπλοκο καθώς εμπλέκονται έννοιες αναγνώρισης προτύπων (pattern recognition) οι οποίες είναι υπολογιστικά ακριβές και δύσκολα υλοποιήσιμες. Εκμεταλλευόμενοι το γεγονός ότι τα “πρότυπα” που πρέπει να αναζητηθούν, τα οποία είναι τα στίγματα φωτεινότητας, ουσιαστικά είναι κύκλοι συγκεκριμένης ακτίνας, υλοποιήθηκε ένας αλγόριθμος ταιριάσματος προτύπων (template matching) για την εύρεση των κέντρων τους. Με γνωστή την ακτίνα των στιγμάτων αυτών, η οποία εκτιμήθηκε σε προηγούμενο βήμα, ο αλγόριθμος αυτός είναι εύκολα υλοποιήσιμος και πιο γρήγορος από τις κλασσικές τεχνικές αναγνώρισης προτύπων.

Η ιδέα του αλγορίθμου αυτού είναι η σάρωση της εικόνας με ένα πρότυπο R , διαστάσεων h_R και w_R (βλ. Σχήμα 4.8.), το οποίο μετατοπίζεται πάνω στην εικόνα αναζήτησης (search image I), η οποία στην προκειμένη περίπτωση είναι η εικόνα DNA, και με χρήση διάφορων “μετρικών απόστασης” υπολογίζεται ένας πίνακας ομοιότητας S (Similarity Matrix).



$$R_{r,s}(u, v) = R(u - r, v - s)$$

Σχήμα 4.8.: Διαδικασία ταιριάσματος προτύπων

Οι μετρικές απόστασης που χρησιμοποιούνται από τον αλγόριθμο, είναι οι εξής:

- Μέγιστη Απόλυτη Διαφορά – Maximum Absolute Difference (mxd)

$$d_M(r, s) = \max_{(i,j) \in R} |I(r + i, s + j) - R(i, j)| \quad (4.7)$$

- Άθροισμα Απολύτων Διαφορών – Sum of Absolute Differences (sad)

$$d_A(r, s) = \sum_{(i,j) \in R} |I(r + i, s + j) - R(i, j)| \quad (4.8)$$

- Άθροισμα Τετραγώνων Διαφορών – Sum of Squared Differences (ssd)

$$d_E^2(r, s) = \sum_{(i,j) \in R} (I(r + i, s + j) - R(i, j))^2 \quad (4.9)$$

- Συσχέτιση – Cross Correlation (ccr)

$$C(r, s) = \sum_{(i,j) \in R} I(r + i, s + j) \cdot R(i, j) \quad (4.10)$$

- Κανονικοποιημένη συσχέτιση – Normalized Cross Correlation (ncc)

$$C_N(r, s) = \frac{\sum_{(i,j) \in R} I(r + i, s + j) \cdot R(i, j)}{\sqrt{\sum_{(i,j) \in R} I(r + i, s + j)} \cdot \sqrt{\sum_{(i,j) \in R} R(i, j)}} \quad (4.11)$$

- Συντελεστές Συσχέτισης – Cross Correlation Coefficients (ccf)

$$C_L(r, s) = \frac{\sum_{(i,j) \in R} (I(r + i, s + j) - \bar{I}(r, s)) \cdot (R(i, j) - \bar{R})}{\sqrt{\sum_{(i,j) \in R} (I(r + i, s + j) - \bar{I}(r, s))^2} \cdot \sqrt{\sum_{(i,j) \in R} (R(i, j) - \bar{R})^2}} \quad (4.12)$$

Ο πίνακας ομοιότητας (S) είναι μια εικόνα διαβαθμίσεων του γκρι, η οποία σε κάθε σημείο της, έχει μια τιμή φωτεινότητας που αντιπροσωπεύει ποσοτικά το “ταίριασμα” του σημείου εκείνου της εικόνας, με το πρότυπο R. Επομένως, σημεία στον πίνακα ομοιότητας με μεγάλες τιμές φωτεινότητας, δηλώνουν πιθανή εμφάνιση του προτύπου R.

Στην προτεινόμενη μεθοδολογία, η μετρική απόστασης που χρησιμοποιείται είναι η μέθοδος κανονικοποιημένης συσχέτισης “ncc” (Σχέση 4.11). Όπως φαίνεται και στα πιο κάτω σχήματα (Σχήματα 4.10.α-ζ), η “ncc” μαζί με την τελευταία μέθοδο “ccf”, δίνουν τα καλύτερα αποτελέσματα, αλλά η “ncc” είναι λιγότερο πολύπλοκη από την “ccf”, γεγονός που την καθιστά καλύτερη επιλογή. Στο Σχήμα 4.9.α παρουσιάζεται το πρότυπο προς αναγνώριση που ζητείται να ανιχνευτεί στην εικόνα αναζήτησης (Σχήμα 4.9.β).



Σχήμα 4.9.α: Πρότυπο προς αναγνώριση



Σχήμα 4.9.β: Εικόνα αναζήτησης

Στα Σχήματα 4.10.α-ζ, φαίνονται τα αποτελέσματα του αλγορίθμου ταιριάσματος προτύπων με χρήση των έξι μετρικών απόστασης αντίστοιχα. Στο επάνω μέρος της κάθε εικόνας παρουσιάζεται ο πίνακας ομοιότητας S και ακριβώς από κάτω παρουσιάζεται η εικόνα αναζήτησης με τα σημεία που ανιχνεύτηκε το πρότυπο R .



Σχήμα 4.10.α: Ταίριασμα εικόνας με χρήση της μετρικής "mxd"



Σχήμα 4.10.β: Ταίριασμα εικόνας με χρήση της μετρικής "sad"



Σχήμα 4.10.γ: Ταίριασμα εικόνας με χρήση της μετρικής "ssd"



Σχήμα 4.10.δ: Ταίριασμα εικόνας με χρήση της μετρικής "ccr"



Σχήμα 4.10.ε: Ταίριασμα εικόνας με χρήση της μετρικής "ncc"

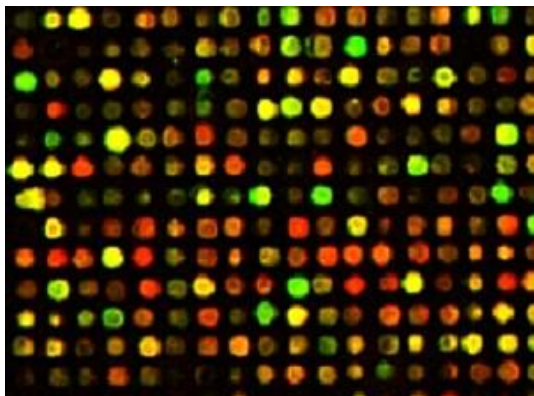


Σχήμα 4.10.ζ: Ταίριασμα εικόνας με χρήση της μετρικής "ccf"

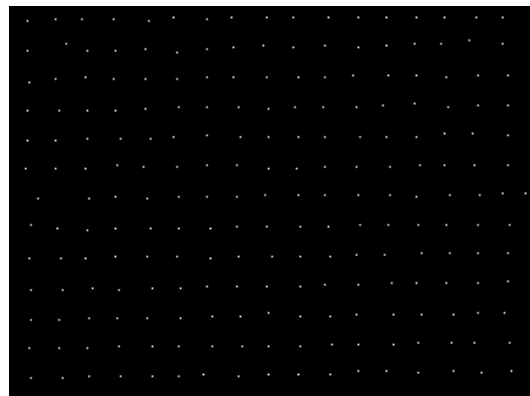
Είναι προφανές ότι οι δύο τελευταίες μέθοδοι, "ncc" και "ccf" δίνουν τα καλύτερα αποτελέσματα, επομένως το μόνο κριτήριο που απομένει για την επιλογή της καλύτερης μεθόδου είναι η ταχύτητα. Λόγω της μικρότερης πολυπλοκότητας της, η "ncc" αποτελεί τη καλύτερη επιλογή.

Για αναζήτηση των σημείων ψηλής φωτεινότητας στον πίνακα ομοιότητας S , εφαρμόζεται ένας αλγόριθμος αναζήτησης τοπικών μεγίστων. Ο αλγόριθμος αυτός, εκτός από τον πίνακα S , παίρνει ως είσοδο και την ακτίνα της περιοχής αναζήτησης των μεγίστων, η οποία επιλέχθηκε να είναι λίγο μεγαλύτερη από την εκτιμώμενη ακτίνα των σιγμάτων, αφού τα κέντρα αυτών είναι που επιδιώκουμε να βρούμε. Μικρότερη ακτίνα θα δώσει περισσότερα τοπικά μέγιστα απ' ότι πρέπει, ενώ αντίθετα, μεγαλύτερη ακτίνα θα αγνοήσει σημαντικά σημεία τοπικών μεγίστων απλά και μόνο επειδή ανήκουν στην περιοχή κάποιου άλλου, ίσως λίγο μεγαλύτερου, τοπικού μεγίστου.

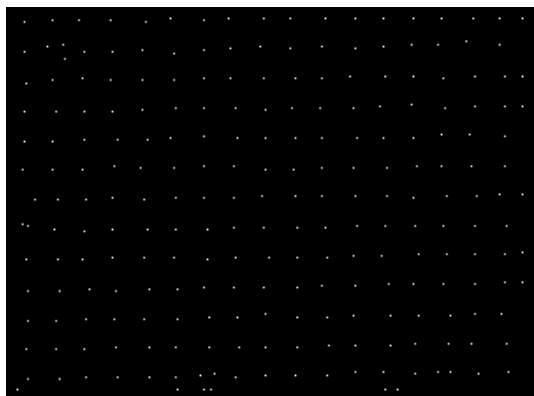
Πιο κάτω παρουσιάζονται αποτελέσματα εφαρμογής του αλγορίθμου εύρεσης κέντρων σιγμάτων στην εικόνα εισόδου (Σχήμα 4.11.α.), με διαφορετικές τιμές ακτινών περιοχής αναζήτησης τοπικών μεγίστων. Με απλή εποπτεία, μπορεί εύκολα να συμπεράνει κανείς ότι πολύ μικρή ακτίνα περιοχής αναζήτησης ανιχνεύει περισσότερα στίγματα από τα πραγματικά (βλ. Σχήμα 4.11.γ στο κάτω μέρος της εικόνας), ενώ αντίστοιχα πολύ μεγάλη ακτίνα (Σχήμα 4.11.δ.) αποτυγχάνει να ανιχνεύσει όλα τα στίγματα στην εικόνα DNA.



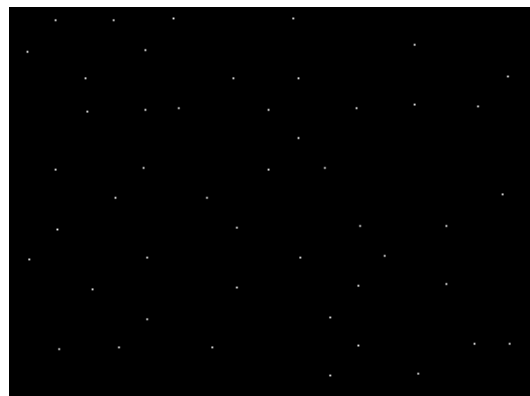
Σχήμα 4.11.α.: Εικόνα εισόδου για ανίχνευση των κέντρων των σιγμάτων



Σχήμα 4.11.β.: Τοπικά μέγιστα με τη κατάλληλη ακτίνα αναζήτησης



Σχήμα 4.11.γ.: Τοπικά μέγιστα με πολύ μικρή ακτίνα αναζήτησης



Σχήμα 4.11.δ.: Τοπικά μέγιστα με πολύ μεγάλη ακτίνα αναζήτησης

Σημειώνεται επίσης ότι λαμβάνονται υπόψη μόνο σημεία που έχουν φωτεινότητα μεγαλύτερη του μισού της μέγιστης φωτεινότητας στην εικόνα. Ο λόγος είναι ότι η φωτεινότητα των σημείων στην προκειμένη περίπτωση, δηλώνει την τιμή της συσχέτισης του προτύπου με την περιοχή της εικόνας αναζήτησης όπου ανιχνεύτηκε το πρότυπο. Κρατώντας μόνο σημεία με συσχέτιση μεγαλύτερη του μισού της μέγιστης που ανιχνεύτηκε, διασφαλίζεται η αξιοπιστία των κέντρων στιγμάτων που ανιχνεύτηκαν.

4.1.5. Υπολογισμός των τετραγωνικών πυραμίδων

Αφού βρεθούν με ακρίβεια τα κέντρα των στιγμάτων, σχεδιάζουμε μια δυαδική εικόνα με τα κέντρα αυτά (η εικόνα αυτή περιέχει μη μηδενικές τιμές, μόνο στα σημεία όπου εκτιμάται ότι βρίσκονται τα κέντρα των στιγμάτων), με διαστάσεις ίσες με αυτές της αρχικής εικόνας DNA. Σαρώνοντας την εικόνα αυτή, σχεδιάζουμε με την βοήθεια μια επαναληπτικής διαδικασίας, ένα-ένα τα επίπεδα των πυραμίδων κεντραρισμένα στις μη μηδενικές τιμές της προηγούμενης δυαδικής εικόνας.

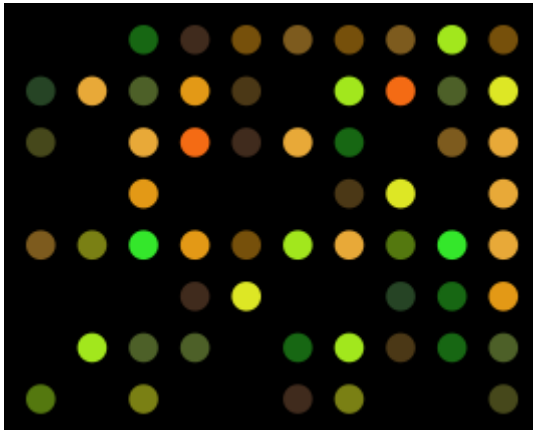
Οι πλευρές των πυραμίδων αυτών, επιλέγονται να είναι λίγο μεγαλύτερες από την εκτιμώμενη ακτίνα των στιγμάτων. Ο λόγος για την επιλογή αυτή, είναι το γεγονός ότι μεταξύ των στιγμάτων, υπάρχει κενό το οποίο πρέπει να ληφθεί υπόψη όταν θέλει κανείς τα χαμηλότερα επίπεδα διαδοχικών πυραμίδων να ενώνονται μεταξύ τους. Τα επίπεδα πρέπει να ενώνονται μεταξύ τους γιατί, όπως θα εξηγήσουμε σε επόμενο στάδιο (βλ. παράγραφο 4.1.7.), οι γραμμές του πλέγματος αναζητούν, σε ένα μέτωπο μπροστά τους, το σημείο με την μικρότερη φωτεινότητα (αντίστοιχα θα μπορούσε να πει κανείς, με το μικρότερο δυναμικό) για να κινηθούν προς αυτό.

Επομένως, η μικρή πλευρά πυραμίδας δεν είναι επιθυμητή επειδή δεν θα υπάρχουν “κουιάδες” μεταξύ των πυραμίδων, επομένως ο αλγόριθμος δεν θα μπορεί να ψάξει για τοπικό ελάχιστο ώστε να προχωρήσουν οι γραμμές του πλέγματος σωστά.

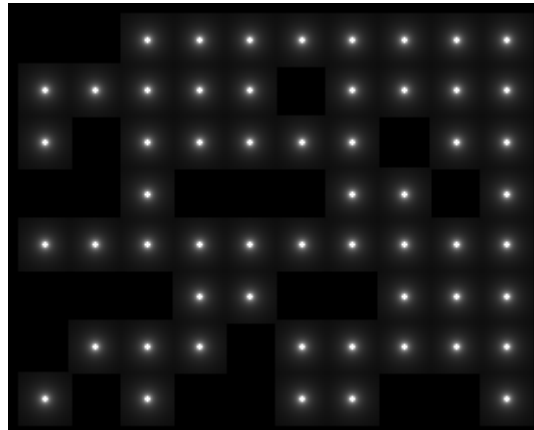
Η ένωση των πυραμίδων είναι προφανώς επιθυμητή, αλλά από την άλλη, πρέπει να υπάρχει και ένα ανώτατο όριο για την πλευρά τους, δηλ. να μην ξεπερνά μια συγκεκριμένη τιμή. Αυτό είναι αναγκαίο, επειδή αν για οποιοδήποτε λόγο δεν ανιχνευθεί κάποιο στίγμα, είτε επειδή δεν υπήρχε εξ' αρχής, είτε επειδή δεν πέτυχε ο αλγόριθμος εύρεσης κέντρων στιγμάτων να το ανιχνεύσει, η προτεινόμενη μεθοδολογία εξακολουθεί να προσπαθεί να εφαρμόσει γραμμές πλέγματος γύρω από το συγκεκριμένο σημείο, ακόμη και αν εκεί δεν έχει σχεδιαστεί πυραμίδα.

Πιο κάτω παρουσιάζονται αποτελέσματα εκτύπωσης πυραμίδων γύρω από τα κέντρα των στιγμάτων που ανιχνεύτηκαν σε μια πραγματική εικόνα DNA (Σχήμα 4.12.α), με τρεις διαφορετικές τιμές πλευράς, για σύγκριση των αποτελεσμάτων. Η εικόνα εισόδου όπως φαίνεται στο Σχήμα 4.12.α, περιέχει κάποια κενά στίγματα για να φανεί η σημασία της ένωσης των πυραμίδων αλλά και ο περιορισμός της πλευράς τους, για να μην καλυφτούν τα κενά στίγματα, ώστε σε επόμενο στάδιο να μπορεί να εφαρμοστεί το πλέγμα γύρω από αυτά. Ο αλγόριθμος σχεδίασης του πλέγματος

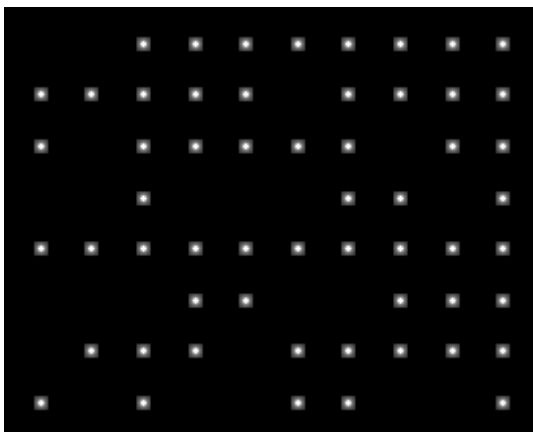
ψάχνει σε ένα μέτωπο μπροστά του για το σημείο με την ελάχιστη φωτεινότητα, ώστε να προχωρήσει η αντίστοιχη γραμμή του πλέγματος προς αυτό το σημείο. Στην περίπτωση του Σχήματος 4.12.δ, ο αλγόριθμος στην προσπάθειά του να αναζητήσει το ελάχιστης φωτεινότητας σημείο για να προχωρήσει, θα αγνοήσει τα κενά και τα κελιά που αντιστοιχούν στα κενά θα είναι υπερβολικά μικρά ίσως και με καθόλου εμβαδόν. Αντιθέτως, στην εικόνα του Σχήματος 4.12.β, ο αλγόριθμος θα εφαρμόσει ένα πλέγμα ώστε να μπορεί ο ερευνητής να αναφέρεται ακόμη και στα κενά στίγματα των οποίων οι θέσεις δεν παύουν να έχουν εξίσου σημαντικό βιολογικό ενδιαφέρον.



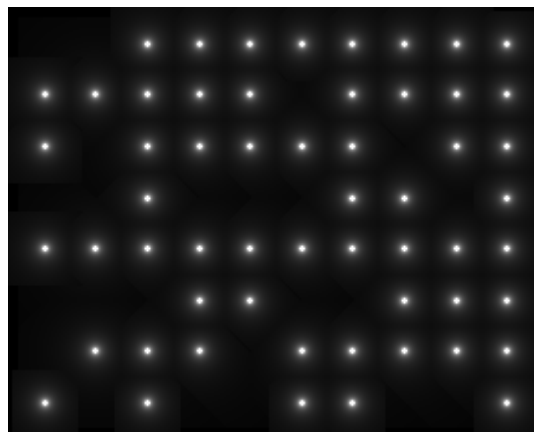
Σχήμα 4.12.α.: Εικόνα εισόδου



Σχήμα 4.12.β.: Πυραμίδες με κατάλληλη πλευρά



Σχήμα 4.12.γ.: Πυραμίδες με πολύ μικρή πλευρά



Σχήμα 4.12.δ.: Πυραμίδες με πολύ μεγάλη πλευρά

4.1.6. Διαδικασία Εφαρμογής του Πλέγματος (gridding process)

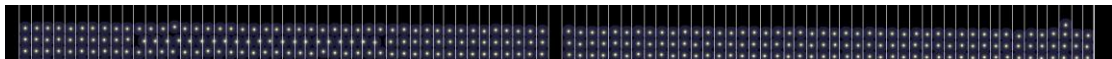
Η σωστή λειτουργία της αυτόματης μεθοδολογίας εφαρμογής πλέγματος που προτείνεται στα πλαίσια της διπλωματικής εργασίας, προϋποθέτει την σωστή επιλογή των αρχικών σημείων από τα οποία θα ξεκινήσουν οι γραμμές του πλέγματος.

Επομένως, σε πρώτο στάδιο, ανιχνεύονται τα “μέτωπα” (“fronts”) που φαίνονται από αριστερά και από πάνω στην εικόνα με τις πυραμίδες. Αυτά τα μέτωπα είναι που θα

ορίσουν τα αρχικά σημεία απ' όπου θα ξεκινήσουν οι οριζόντιες και κατακόρυφες γραμμές του πλέγματος αντίστοιχα.

Αφού ανιχνευτούν τα μέτωπα αυτά, επιλέγονται σημεία όπως για παράδειγμα αυτά που έχουν διαφορετικό ύψος στα μέτωπα που ανιχνεύτηκαν στο προηγούμενο στάδιο (διαφορετική τετμημένη για το κατακόρυφο μέτωπο που φαίνεται από αριστερά και διαφορετική τεταγμένη για το οριζόντιο μέτωπο που φαίνεται από πάνω), αλλά και σημεία των μετώπων τα οποία αντιστοιχούν σε τοπικά ελάχιστα στη εικόνα με τις πυραμίδες. Επίσης επιλέγονται και σημεία όπου παρουσιάζονται αλλαγές από 0 σε 1 και αντίστροφα από 1 σε 0, στις κατωφλιοποιημένες προβολές που χρησιμοποιούνται για ανίχνευση των υποεικόνων. Αυτά τα σημεία, τα οποία αντιστοιχούν στις ακμές των υποεικόνων, είναι πολύ σημαντικά στις περιπτώσεις που ο ερευνητής επιθυμεί να εφαρμόσει ένα πλέγμα σε ολόκληρη την εικόνα DNA (με όλες τις υποεικόνες) ενώ ταυτόχρονα απαιτείται και η παράλληλη ανίχνευση των υποεικόνων που περιέχει η εικόνα DNA. Σε επόμενο στάδιο, συμπληρώνονται τυχόν μεγάλα κενά μεταξύ των σημείων που επιλέχθηκαν.

Στα επόμενα σχήματα (Σχήματα 4.13.α και 4.13.β), παρουσιάζονται τα “μέτωπα” που ανιχνεύθηκαν, σε μια εικόνα DNA διαστάσεων 1800x1800 pixels, που περιέχει τέσσερις (4) υποεικόνες, μαζί με τα σημεία που επιλέχθηκαν για αρχικοποίηση του πλέγματος. Τα σημεία αυτά παρουσιάζονται με ευθείς γραμμές που διατρέχουν τα αντίστοιχα σημεία πάνω στο εκάστοτε “μέτωπο”.



Σχήμα 4.13.α.: Οριζόντιο “μέτωπο” εικόνας πυραμίδων



Σχήμα 4.13.β.: Κατακόρυφο “μέτωπο” εικόνας πυραμίδων

Με υπολογισμένα τα αρχικά σημεία για τις γραμμές του πλέγματος, η δουλειά του αλγορίθμου σχεδιασμού του πλέγματος είναι πλέον πολύ εύκολη. Καθώς διατρέχεται ολόκληρη η εικόνα DNA pixel προς pixel, εφαρμόζονται στο εκάστοτε εικονοστοιχείο, δύο συναρτήσεις (μια φορά για τις κατακόρυφες και μια φορά για τις οριζόντιες γραμμές του πλέγματος). Η δουλειά αυτών των συναρτήσεων είναι η αναζήτηση των σημείων ελάχιστης φωτεινότητας σε ένα μέτωπο, μπροστά από το εικονοστοιχείο αυτό, 3 ή και περισσότερων αν επιθυμείται, εικονοστοιχείων. Με αυτό το τρόπο προχωράει η αντίστοιχη γραμμή του πλέγματος, μέχρι το τέλος της εικόνας DNA.

4.1.7. Αποθήκευση των Δεδομένων του Πλέγματος

Οι συντεταγμένες του κάθε εικονοστοιχείου (pixel) του πλέγματος αποθηκεύονται ώστε να μην χρειάζεται να εφαρμόζεται κάθε φορά η όλη διαδικασία από την αρχή. Λόγω της ιδιαίτερης μορφής του πλέγματος δεν είναι αρκετή η πληροφορία μόνο των αρχικών σημείων στον οριζόντιο και κατακόρυφο άξονα και η γωνία περιστροφής της εικόνας, για τον πλήρη επανασχεδιασμό του πλέγματος. Αν και σε άλλες μεθόδους, όπως για παράδειγμα της μεθόδου προβολών, τα τρία αυτά στοιχεία είναι υπεραρκετά για την πλήρη περιγραφή του πλέγματος, στην περίπτωση της προτεινόμενης μεθοδολογίας χρειάζεται περισσότερη πληροφορία.

Συγκεκριμένα, αποθηκεύονται δύο πίνακες. Ένας για τις κατακόρυφες γραμμές του πλέγματος (πλήθους M) και ένας για τις οριζόντιες (πλήθους N), όπως φαίνεται στα Σχήματα 4.14.α και 4.14.β. Ο πρώτος πίνακας (xGrid) έχει τόσες γραμμές όσα είναι και τα εικονοστοιχεία που έχει η πρώτη (η κατακόρυφη) διάσταση της εικόνας (h). Οι στήλες του αντιπροσωπεύουν τις κατακόρυφες γραμμές του πλέγματος (πλήθους M) και κάθε στοιχείο του πίνακα (μπλε κελί) δηλώνει την απόλυτη τεταγμένη του συγκεκριμένου εικονοστοιχείου της εκάστοτε κατακόρυφης γραμμής του πλέγματος. Αντίστοιχα, ο δεύτερος πίνακας (yGrid) αποτελείται από τόσες στήλες όσες και τα εικονοστοιχεία της δεύτερης (της οριζόντιας) διάστασης της εικόνας (w) και γραμμές όσες οι οριζόντιες γραμμές του πλέγματος (πλήθους N). Κάθε στοιχείο του δεύτερου πίνακα (κόκκινο κελί) δηλώνει την απόλυτη τεταγμένη του συγκεκριμένου εικονοστοιχείου της εκάστοτε οριζόντιας γραμμής του πλέγματος.

	1	2	3	4		M
1						
2						
3						
4						
h						

Σχήμα 4.14.α.: Πίνακας αποθήκευσης
 $xGrid(i,jCell)$

	1	2	3	4		w
1						
2						
3						
4						
N						

Σχήμα 4.14.β.: Πίνακας αποθήκευσης
 $yGrid(iCell,j)$

Με αυτό τον τρόπο, μπορεί εύκολα στο μέλλον ο ερευνητής, διατρέχοντας το πλέγμα κελί προς κελί, να το εκτυπώνει πάνω σε μια εικόνα DNA και να εμφανίζεται το ίδιο αποτέλεσμα που τυπώθηκε όταν εφαρμόστηκε για πρώτη φορά η προτεινόμενη αυτόματη μεθοδολογία.

4.1.8. Διαδικασία βελτίωσης του πλέγματος (Refinement procedure)

Για πιο αξιόπιστα αποτελέσματα, στο τέλος της εφαρμογής του πλέγματος, εφαρμόζεται μια διαδικασία βελτίωσης του ώστε να διορθωθούν τυχόν ατέλειες που μπορεί να παρουσιαστούν.

Η διαδικασία βελτίωσης, αποτελείται από πέντε στάδια: α) επιβεβαίωση ύπαρξης των οριακών γραμμών (boundaries) στο πλέγμα, β) διαγραφή των περισσών γραμμών, γ) διαχωρισμός κελιών που περιέχουν περισσότερα από ένα στίγματα, δ) σπρώξιμο των οριακών γραμμών του πλέγματος για να πλησιάσουν τα ακραία στίγματα και ε) ένα γενικό ραφινάρισμα των συντεταγμένων των γραμμών που αποτελούν το πλέγμα για τη βέλτιστη εφαρμογή του στην εικόνα DNA.

Επειδή η επιλογή των αρχικών σημείων των γραμμών του πλέγματος, εξαρτάται σημαντικά κατά ένα μέρος από την σωστή ανίχνευση των υποεικόνων, οι οριακές γραμμές του πλέγματος πολύ πιθανών να μην εμφανίζονται στο πλέγμα για τον λόγο ότι μπορεί να ανήκουν σε περιοχή που δεν περιλαμβάνεται σε μια από τις υποεικόνες που ανιχνεύτηκαν. Η λύση στο πρόβλημα αυτό είναι η αναζήτηση κέντρων στιγμάτων που πέφτουν έξω από τις οριακές γραμμές του πλέγματος. Αν ανιχνευθεί έστω και ένα κέντρο στίγματος, το οποίο να μην περικλείεται και από την εξωτερική πλευρά από γραμμή του πλέγματος, τότε προστίθεται μια επιπλέον γραμμή στο πλέγμα για την κάλυψη αυτού του στίγματος. Εκμεταλλευόμενος το γεγονός ότι η εικόνα με τα κέντρα των στιγμάτων είναι δυαδικής μορφής (binary), με αθροίσματα σειρών και στις δύο διαστάσεις συγκεκριμένων περιοχών στην εικόνα (Σχέση 4.13), μπορεί κανείς να μετρήσει πόσα κέντρα περιέχει η εκάστοτε περιοχή. Η περιοχή αναζήτησης στιγμάτων πάνω στην εικόνα περιορίζεται από το αριστερότερο όριο $wStart$ μέχρι και το δεξιότερο όριο $wStop$ και από το ανώτερο όριο $hStart$ μέχρι και το κατώτερο όριο $hStop$.

$$NumOfSpots = \sum_{j=wStart}^{wStop} \sum_{i=hStart}^{hStop} I(i,j) \quad (4.13)$$

Με αυτό τον τρόπο αναζητούνται κέντρα στιγμάτων που απέτυχε ο αλγόριθμος του προηγούμενου βήματος να τα καλύψει.

Στο επόμενο βήμα βελτίωσης του πλέγματος, ακολουθώντας την ίδια διαδικασία με πιο πάνω, αναζητούνται κελιά που δεν περιέχουν κέντρα στιγμάτων. Στην πραγματικότητα αναζητούνται κελιά τα οποία στα τέσσερα πέμπτα (4/5) του συνόλου τους, στην ίδια γραμμή ή στήλη, δεν περιέχουν κέντρα στιγμάτων. Ο λόγος που δεν τα αναζητούμε για όλο το σύνολο της γραμμής ή στήλης, είναι επειδή υπάρχουν και περιπτώσεις που μπορεί κάποια κελιά απλά να χρειάζονται βελτίωση των γραμμών που τα περικλείουν (η επίλυση τέτοιων προβλημάτων είναι ευθύνη του τελευταίου σταδίου της διαδικασίας βελτίωσης). Αν βρεθεί μια τέτοια περίπτωση στο i -οστό κελί, τότε διαγράφεται η i -οστή γραμμή του πλέγματος αφού σύμφωνα με τον

τρόπο σχεδιασμού του πλέγματος, το i -οστό κελί βρίσκεται μεταξύ των διαδοχικών i και $i+1$ γραμμών του πλέγματος. Αυτό ισχύει τόσο στις κατακόρυφες όσο και στις οριζόντιες γραμμές του (όπου ο δείκτης i αντιστοιχεί στο δείκτη j).

Αντίστοιχα με το προηγούμενο βήμα, αναζητούνται κελιά που περιέχουν περισσότερα από δύο κέντρα στιγμάτων, σύμφωνα με την Σχέση 4.13. Αν το άθροισμα, στην περιοχή του συγκεκριμένου κελιού, είναι μεγαλύτερο από ένα, σημαίνει ότι περιέχονται 2 ή και περισσότερα κέντρα στιγμάτων. Και σε αυτή την περίπτωση, αναζητούνται κελιά τα οποία στα τέσσερα πέμπτα (4/5) του συνόλου τους στην ίδια γραμμή ή στήλη, δεν περιέχουν κέντρα στιγμάτων. Ο λόγος είναι, όπως πιο πάνω, το γεγονός ότι υπάρχουν και περιπτώσεις που μπορεί τα κελιά αυτά να χρειάζονται απλά βελτίωση των γραμμών που τα περικλείουν. Στην περίπτωση αυτή, αν ανιχνευτούν πολλαπλά στίγματα στο i -οστό κελί, προστίθεται μια επιπλέον γραμμή μετά την i -οστή γραμμή του πλέγματος. Οι θέσεις των εικονοστοιχείων της καινούριας εμβόλιμης γραμμής, υπολογίζονται έτσι ώστε να τοποθετηθεί ενδιάμεσα των εκατέρωθεν γραμμών της, σύμφωνα με τις Σχέσεις 4.14.α και 4.14.β. Επειδή υπάρχει διαίρεση στην πράξη αυτή και επειδή οι συντεταγμένες των εικονοστοιχείων πρέπει να είναι ακέραιοι αριθμοί, είναι αναγκαία η στρογγυλοποίηση σε ακέραιο αριθμό. Αυτό επιτυγχάνεται με την συνάρτηση $round()$.

$$xGridNew(i,jCell) = round((xGridOld(i,jCell) + xGridOld(i,jCell + 1))/2) \quad (4.14.α)$$

$$yGridNew(iCell,j) = round((yGridOld(iCell,j) + yGridOld(iCell + 1,j))/2) \quad (4.14.β)$$

Αν οι θέσεις της γραμμής αυτής χρειαστούν κάποια ρύθμιση, αυτό θα γίνει στο τελευταίο στάδιο της διαδικασίας βελτίωσης των γραμμών του πλέγματος.

Ειδικά στις περιπτώσεις εφαρμογής του πλέγματος σε εικόνες που έχουν παραμορφωθεί από ημιτονοειδείς μετασχηματισμούς, οι οριακές, τόσο οι οριζόντιες όσο και οι κατακόρυφες, γραμμές του πλέγματος απαιτούν επιπλέον ρύθμιση για να πλησιάσουν τα ακραία στίγματα. Ο λόγος που καθιστά αναγκαία την επιπλέον αυτή ρύθμιση, είναι το γεγονός ότι δεν υπάρχει η αναγκαία πληροφορία στις ακραίες θέσεις του πλέγματος έτσι ώστε να εφαρμοστούν σωστά οι οριακές γραμμές (στην περίπτωση της προτεινόμενης μεθοδολογίας, αυτή η πληροφορία προσφέρεται από τις πυραμίδες). Αν και αυτό το στάδιο δεν είναι καίριο στην όλη μεθοδολογία, είναι σημαντικό σε περιπτώσεις που ο ερευνητής αναφέρεται σε συγκεκριμένο στίγμα και επιθυμεί να παρουσιάζεται η εγγύτερη, σε αυτό το στίγμα, περιοχή και όχι μια μεγαλύτερη, για να μπορεί να εξάγει πιο ασφαλή και αξιόπιστα συμπεράσματα. Η διαδικασία που ακολουθεί το στάδιο αυτό θα μπορούσε να παρομοιαστεί με μια “πρέσα”, η οποία “σπρώχνει” τις ακραίες γραμμές του πλέγματος (κατακόρυφες και οριζόντιες) προς το κέντρο της εικόνας, υπό την αυστηρή ικανοποίηση συγκεκριμένων περιορισμών, όπως για παράδειγμα να μην διασταυρώσουν

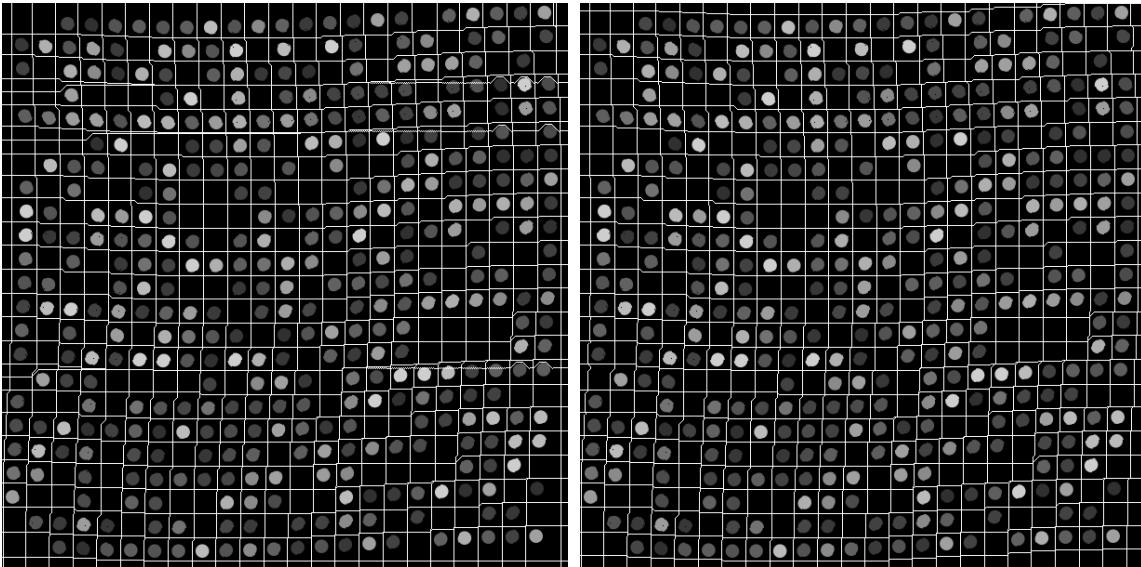
υπάρχουσες γραμμές του πλέγματος, για την εξασφάλιση της σωστής εφαρμογής του στην εικόνα DNA.

Το τελικό στάδιο του ραφινάριατος (refinement) των γραμμών του πλέγματος είναι το πλέον σημαντικότερο βήμα της διαδικασίας βελτίωσης. Η οποιαδήποτε εσφαλμένη τοποθέτηση μιας κατακόρυφης ή οριζόντιας γραμμής του πλέγματος, σε αυτό το βήμα θα διορθωθεί με τον κατάλληλο τρόπο για να δώσει το βέλτιστο αποτέλεσμα.

Σαρώνονται όλα τα κελιά του πλέγματος το οποίο εφαρμόστηκε πάνω στην εικόνα DNA και όπως σε προηγούμενο βήμα, ανιχνεύονται τα στίγματα που περιέχει το κάθε κελί, σύμφωνα με τη Σχέση 4.13. Αν βρεθεί κελί το οποίο περιέχει δύο στίγματα, σε πρώτη φάση περνά κάποια επεξεργασία ώστε να αποφασιστεί αν πρέπει να διαχωριστεί σε δύο διαφορετικές στήλες ή γραμμές. Ένας γρήγορος τρόπος είναι να μετρηθούν οι διαστάσεις του κελιού και αν η πρώτη (η κατακόρυφη) διάσταση είναι μεγαλύτερη από τη δεύτερη (την οριζόντια) διάσταση τότε το κελί πρέπει να διαχωριστεί σε δύο διαφορετικές γραμμές με μια επιπλέον ενδιάμεση οριζόντια γραμμή, ενώ στην αντίθετη περίπτωση το κελί πρέπει να διαχωριστεί σε δύο διαφορετικές στήλες με μια επιπλέον ενδιάμεση κάθετη γραμμή.

Αφού αποφασιστεί ο τρόπος διαχωρισμού του κελιού, η μεθοδολογία προχωρά στην ρύθμιση των γραμμών του πλέγματος. Αν για παράδειγμα το i -οστό κελί πρέπει να διαχωριστεί σε δύο στήλες, τότε η $i+1$ κατακόρυφη γραμμή του πλέγματος μεταφέρεται, μόνο για τις τιμές που αντιστοιχούν στο ύψος του συγκεκριμένου κελιού, στο μέσο του διαστήματος μεταξύ των i και $i+2$ γραμμών του πλέγματος. Αφού μεταφερθεί η $i+1$ κατακόρυφη γραμμή του πλέγματος στην καινούρια θέση της, επαναπροσδιορίζεται η θέση του κάθε εικονοστοιχείου (pixel) της ανάλογα με την εικόνα των πυραμίδων που κατασκευάστηκε σε προηγούμενο στάδιο. Αυτό επιτυγχάνεται ψάχνοντας στην εικόνα με τις πυραμίδες για τα σημεία με την ελάχιστη φωτεινότητα.

Στα δύο επόμενα σχήματα (Σχήματα 4.15.α και 4.15.β) παρουσιάζεται μια περίπτωση εικόνας DNA μετά από εφαρμογή πλέγματος, στην οποία η διαδικασία βελτίωσης που περιγράφεται πιο πάνω επιφέρει πολύ σημαντικές αλλαγές στο πλέγμα και δίνει το βέλτιστο αποτέλεσμα. Αναλυτικότερα, στο Σχήμα 4.15.α, παρουσιάζεται το αποτέλεσμα εφαρμογής πλέγματος στην εικόνα DNA πριν την διαδικασία βελτίωσης και στο Σχήμα 4.15.β, παρουσιάζεται το αποτέλεσμα της εφαρμογής πλέγματος στην ίδια εικόνα DNA μετά από την διαδικασία βελτίωσης.



Σχήμα 4.15.α.: Αποτέλεσμα πριν την διαδικασία βελτίωσης

Σχήμα 4.15.β.: Αποτέλεσμα μετά την διαδικασία βελτίωσης

Με απλή εποπτεία, φαίνεται καθαρά τόσο η αφαίρεση των εσφαλμένων περισσών γραμμών (πάνω αριστερά, πάνω δεξιά και στην μέση αριστερά) καθώς και η ρύθμιση των ακραίων γραμμών του πλέγματος (κάτω δεξιά), ώστε να επιτευχθεί το βέλτιστο αποτέλεσμα.

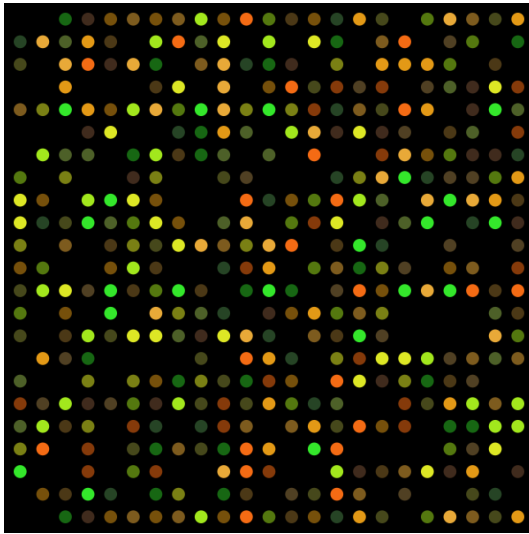
Αποτελέσματα Εφαρμογής Πλέγματος

5.1. Αποτελέσματα Εφαρμογής Πλέγματος σε ιδανικές εικόνες DNA που έχουν παραμορφωθεί από γνωστούς μετασχηματισμούς

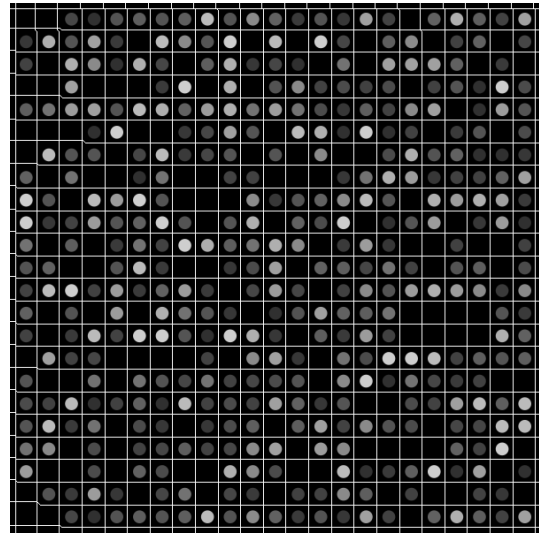
Προκειμένου να αξιολογηθεί η προτεινόμενη αυτόματη μεθοδολογία εφαρμογής πλέγματος, χρησιμοποιούνται δύο σύνολα εικόνων εισόδου. Το πρώτο σύνολο περιλαμβάνει ιδανικές εικόνες DNA που έχουν παραμορφωθεί από γνωστούς μετασχηματισμούς και το δεύτερο σύνολο περιλαμβάνει πραγματικές εικόνες DNA. Τα αποτελέσματα της μεθοδολογίας παρουσιάζονται στις επόμενες παραγράφους.

Η ιδανική εικόνα DNA που φαίνεται στο Σχήμα 5.1.α. είναι η εικόνα εισόδου, διαστάσεων 600x600 pixels, η οποία θα παραμορφωθεί και ακολούθως θα χρησιμοποιηθεί ως είσοδος για εφαρμογή πλέγματος. Στο Σχήμα 5.1.β. παρουσιάζεται το αποτέλεσμα εφαρμογής πλέγματος στην ιδανική εικόνα εισόδου DNA (Σχήμα 5.1.α).

Η εικόνα αναφοράς (Σχήμα 5.1.α) θα παραμορφωθεί σε τέσσερις αντίστοιχες εικόνες, σε πρώτο στάδιο σύμφωνα με δύο αφφινικούς μετασχηματισμούς και αργότερα με δύο ημιτονοειδείς μετασχηματισμούς. Οι παράμετροι των τεσσάρων μετασχηματισμών επιλέχθηκαν με τέτοιο τρόπο ώστε να καλύπτουν ένα ευρύ φάσμα παραμορφώσεων, από σχετικά μικρές παραμορφώσεις (Αφφινικός-1 και Ημιτονοειδής-1), μέχρι και σχετικά μεγάλες (Αφφινικός-2 και Ημιτονοειδής-2).



Σχήμα 5.1.α: Ιδανική εικόνα εισόδου DNA (χωρίς μετασχηματισμό)



Σχήμα 5.1.β: Αποτέλεσμα εφαρμογής πλέγματος σε ιδανική εικόνα DNA

5.1.1. Αφφινικός Μετασχηματισμός

Η συνάρτηση του αφφινικού μετασχηματισμού περιγράφεται ως εξής:

$$X = s \cdot \cos\theta \cdot (x - x_c) - s \cdot \sin\theta \cdot (y - y_c) + t_1 + x_c \quad (5.1.α)$$

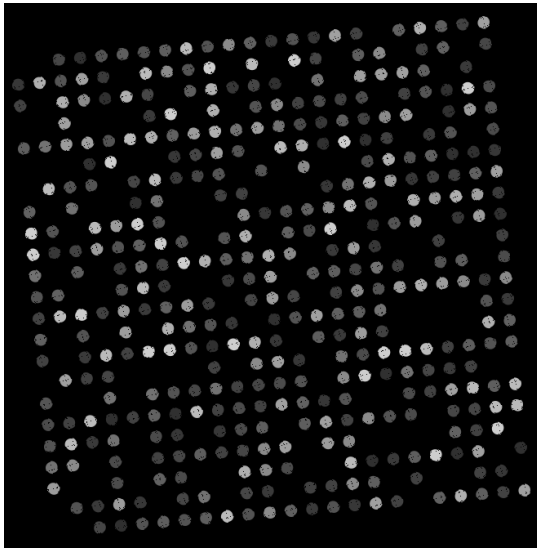
$$Y = s \cdot \sin\theta \cdot (x - x_c) - s \cdot \cos\theta \cdot (y - y_c) + t_2 + y_c \quad (5.1.β)$$

όπου $x(X)$ και $y(Y)$ είναι οι δείκτες της στήλης και της γραμμής, αντίστοιχα, στην αρχική εικόνα αναφοράς, $(x_c, y_c) = (N_x/2, N_y/2)$ το κέντρο της εικόνας και N_x, N_y το πλήθος των και γραμμών, αντίστοιχα, στην εικόνα αναφοράς, s η παράμετρος κλιμάκωσης, θ η γωνία περιστροφής και t_1, t_2 οι μετατοπίσεις στον οριζόντιο και κατακόρυφο άξονα, αντίστοιχα.

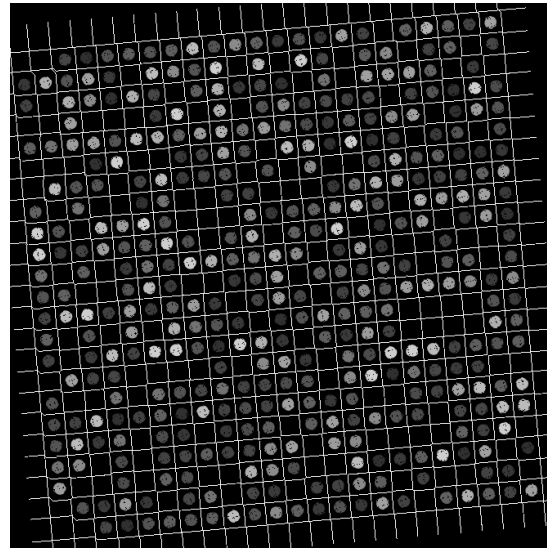
Αφφινικός – 1

Οι παράμετροι που χρησιμοποιήθηκαν είναι:

$$s = 1.02, \theta = 5^\circ, t_1 = 5, t_2 = -5.$$



Σχήμα 5.2.α: Ιδανική εικόνα DNA παραμορφωμένη από Αφφινικό-1

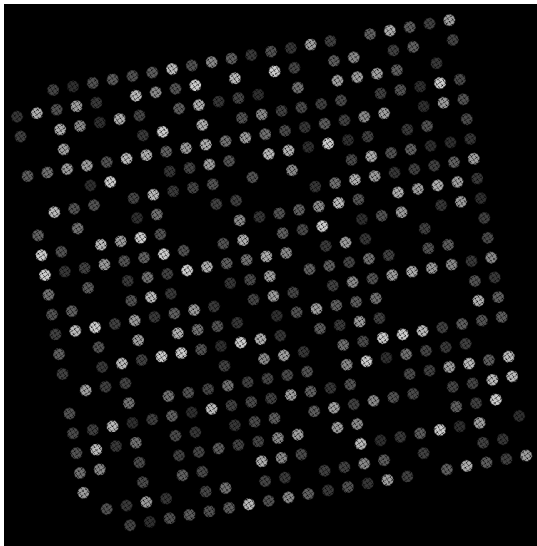


Σχήμα 5.2.β: Αποτέλεσμα εφαρμογής πλέγματος στην εικόνα με Αφφινικό-1

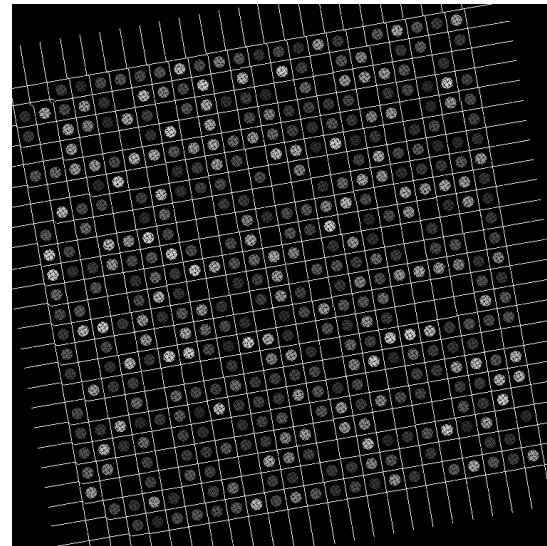
Αφφινικός – 2

Οι παράμετροι που χρησιμοποιήθηκαν είναι:

$$s = 1.1, \vartheta = 10^\circ, t_1 = 10 \quad t_2 = -10.$$



Σχήμα 5.3.α: Ιδανική εικόνα DNA παραμορφωμένη από Αφφινικό-2



Σχήμα 5.3.β: Αποτέλεσμα εφαρμογής πλέγματος στην εικόνα με Αφφινικό-2

5.1.2. Ημιτονοειδής Μετασχηματισμός

Η συνάρτηση του ημιτονοειδή μετασχηματισμού περιγράφεται ως εξής:

$$X = x + a \cdot \sin\left(\frac{y}{T}\right) \quad (5.2.α)$$

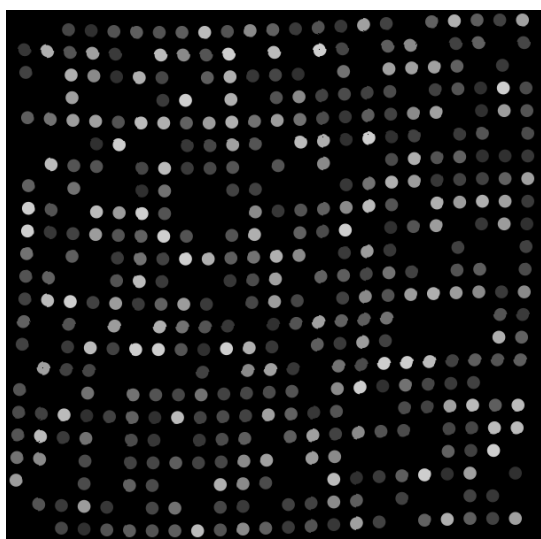
$$Y = y + a \cdot \cos\left(\frac{x}{T}\right) \quad (5.2.β)$$

όπου a το πλάτος του ημιτόνου και T η περίοδος του.

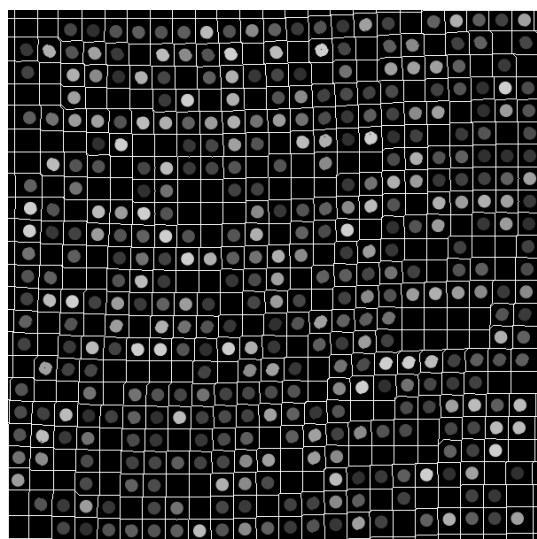
Ημιτονοειδής Μετασχηματισμός – 1

Οι παράμετροι που χρησιμοποιήθηκαν είναι:

$$a = 7, T = 120.$$



Σχήμα 5.4.α: Ιδανική εικόνα DNA παραμορφωμένη από Ημιτονοειδή-1

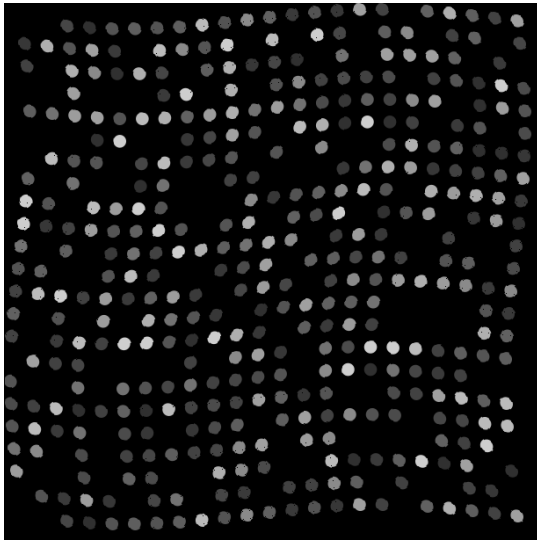


Σχήμα 5.4.β: Αποτέλεσμα εφαρμογής πλέγματος στην εικόνα με Ημιτονοειδή -1

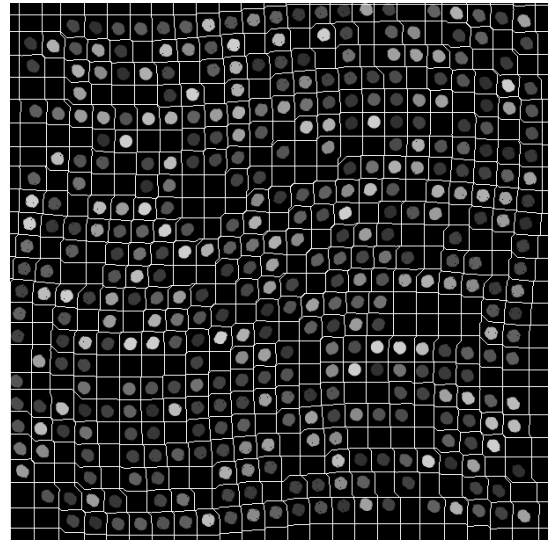
Ημιτονοειδής Μετασχηματισμός – 2

Οι παράμετροι που χρησιμοποιήθηκαν είναι:

$$a = 11, T = 90.$$



Σχήμα 5.5.α: Ίδανική εικόνα DNA παραμορφωμένη από Ημιτονοειδή-2



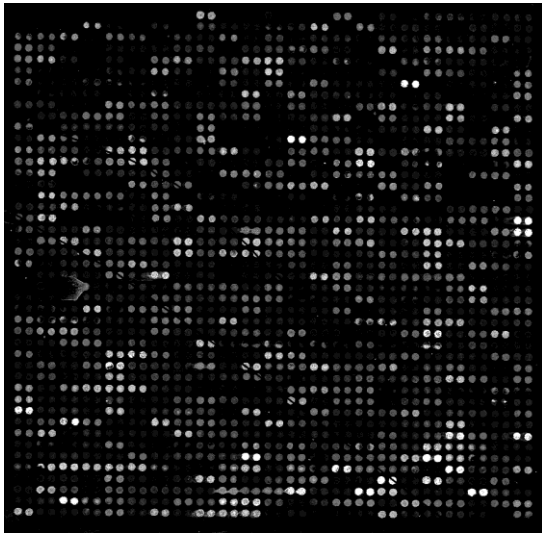
Σχήμα 5.5.β: Αποτέλεσμα εφαρμογής πλέγματος στην εικόνα με Ημιτονοειδή -2

5.2. Αποτελέσματα Εφαρμογής Πλέγματος σε πραγματικές εικόνες DNA (άγνωστος μετασχηματισμός)

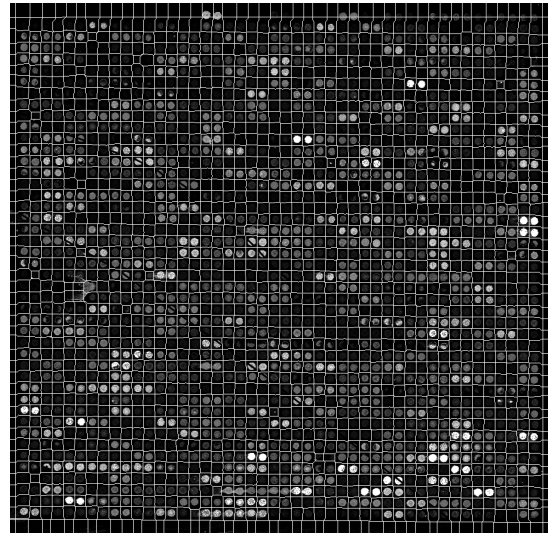
Οι ερευνητές συνήθως δεν αντιμετωπίζουν εικόνες DNA όπως τις προηγούμενες, καθώς δεν είναι συνηθισμένο να συναντήσει κανείς εικόνα παραμορφωμένη από ημιτονοειδή μετασχηματισμό. Προκειμένου να αποδειχθεί η ποιότητα των αποτελεσμάτων της προτεινόμενης μεθοδολογίας σε πραγματικές εικόνες DNA, παρουσιάζονται ενδεικτικά αποτελέσματα εφαρμογής πλέγματος σε υποεικόνες πραγματικών εικόνων DNA.

Στα Σχήματα 5.6.α-γ και 5.7.α-γ, παρουσιάζονται τα αποτελέσματα εφαρμογής πλέγματος χρησιμοποιώντας τη προτεινόμενη αυτόματη μεθοδολογία, στις υποεικόνες (1,2) και (2,2) αντίστοιχα, μιας πραγματικής εικόνας DNA η οποία περιέχει 4 υποεικόνες. Η υποεικόνα (1,2) έχει διαστάσεις 887x905 pixels ενώ η υποεικόνα (2,2) 891x905 pixels.

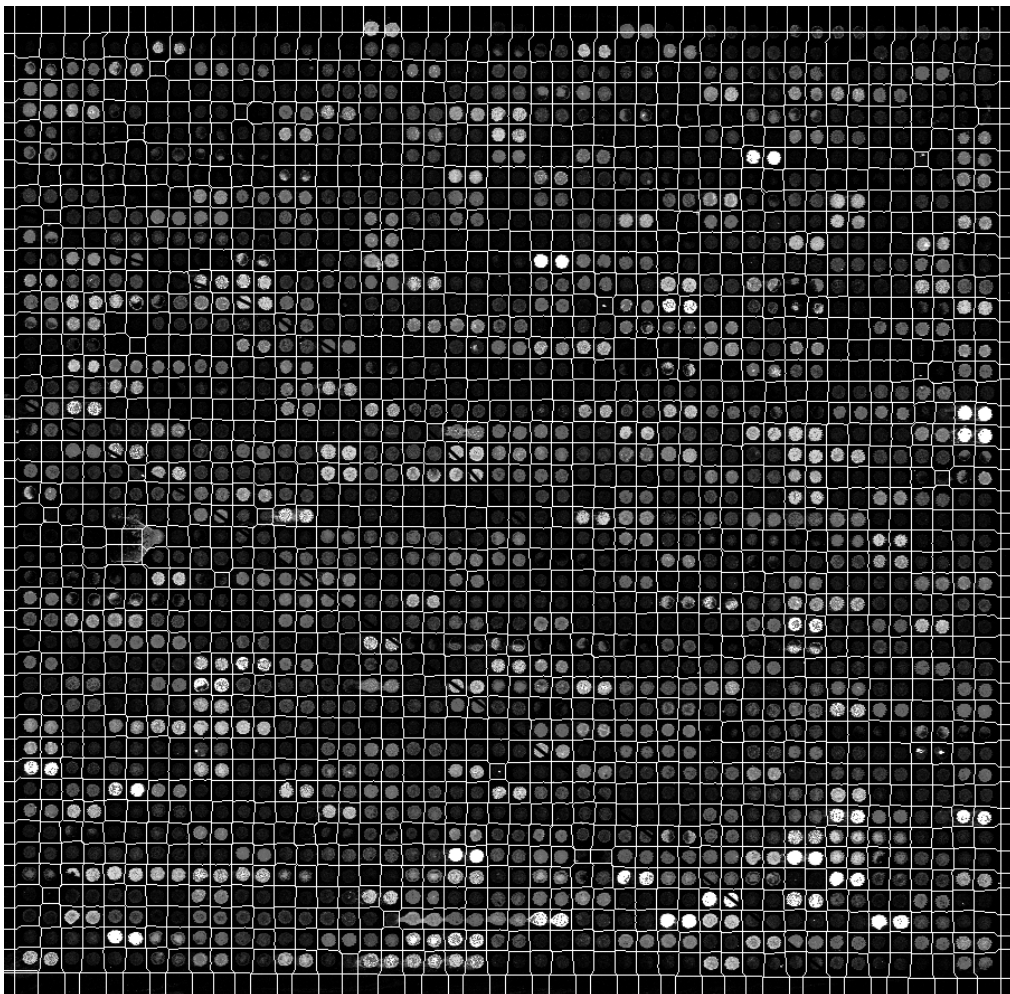
Για καλύτερη εμφάνιση των αποτελεσμάτων, στα πάνω σχήματα παρουσιάζονται η εικόνα εισόδου και δίπλα το αποτέλεσμα εφαρμογής του πλέγματος ενώ ακριβώς από κάτω παρουσιάζεται το αποτέλεσμα εφαρμογής πλέγματος μόνο του σε μεγέθυνση για καλύτερη αξιολόγηση των αποτελεσμάτων.



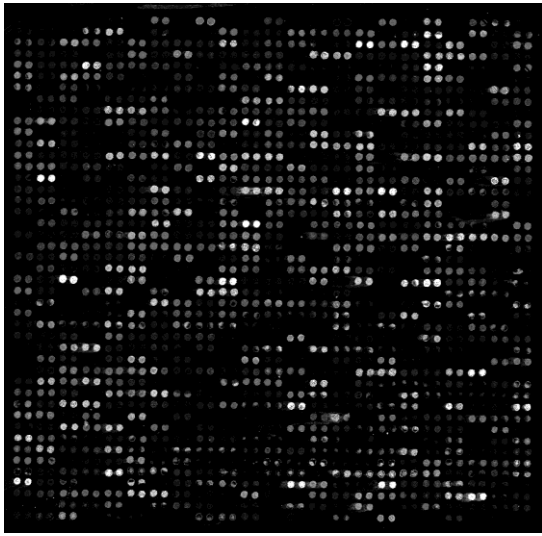
Σχήμα 5.6.α.: Υποεικόνα (1,2) της πραγματικής εικόνας DNA



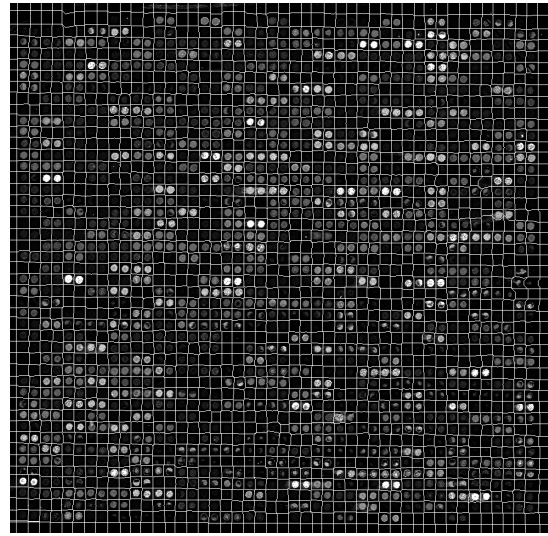
Σχήμα 5.6.β.: Αποτέλεσμα εφαρμογής πλέγματος στην υποεικόνα (1,2)



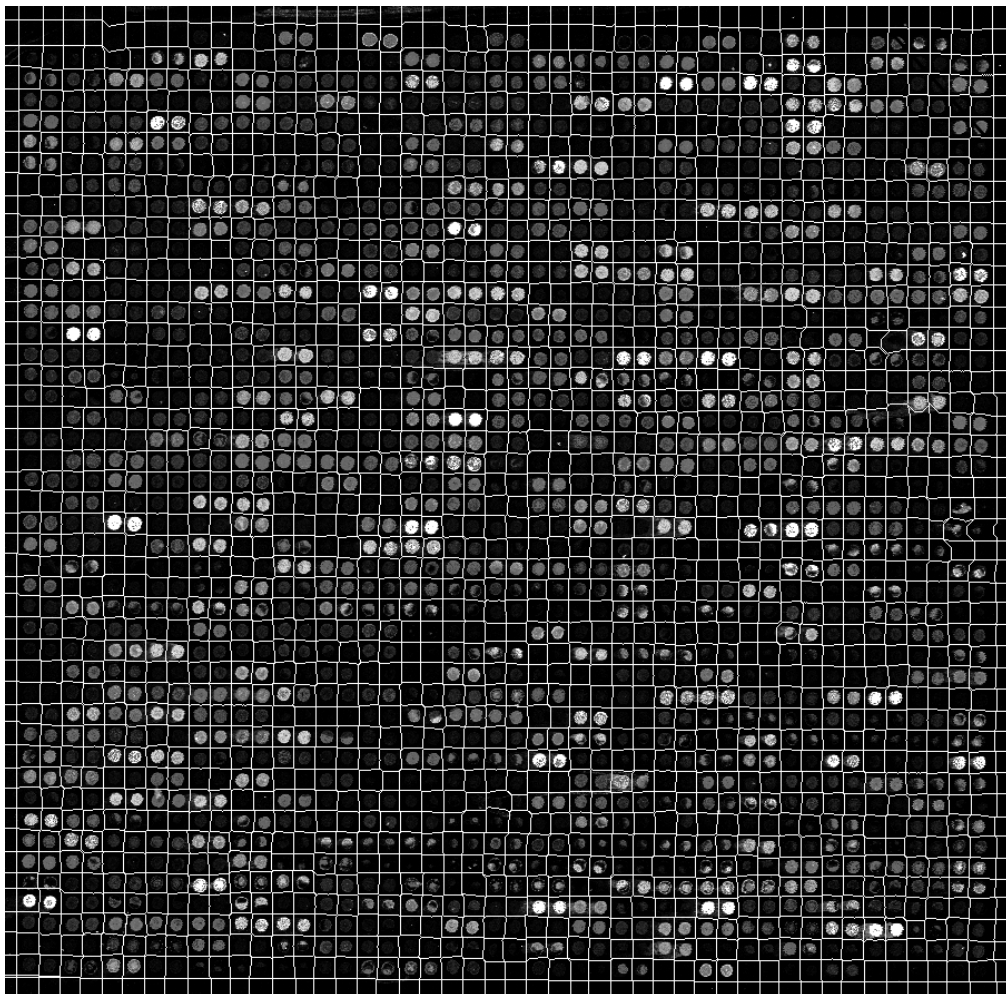
Σχήμα 5.6.γ.: Αποτέλεσμα εφαρμογής πλέγματος στην υποεικόνα (1,2) σε μεγέθυνση



Σχήμα 5.7.α.: Υποεικόνα (2,2) της πραγματικής εικόνας DNA

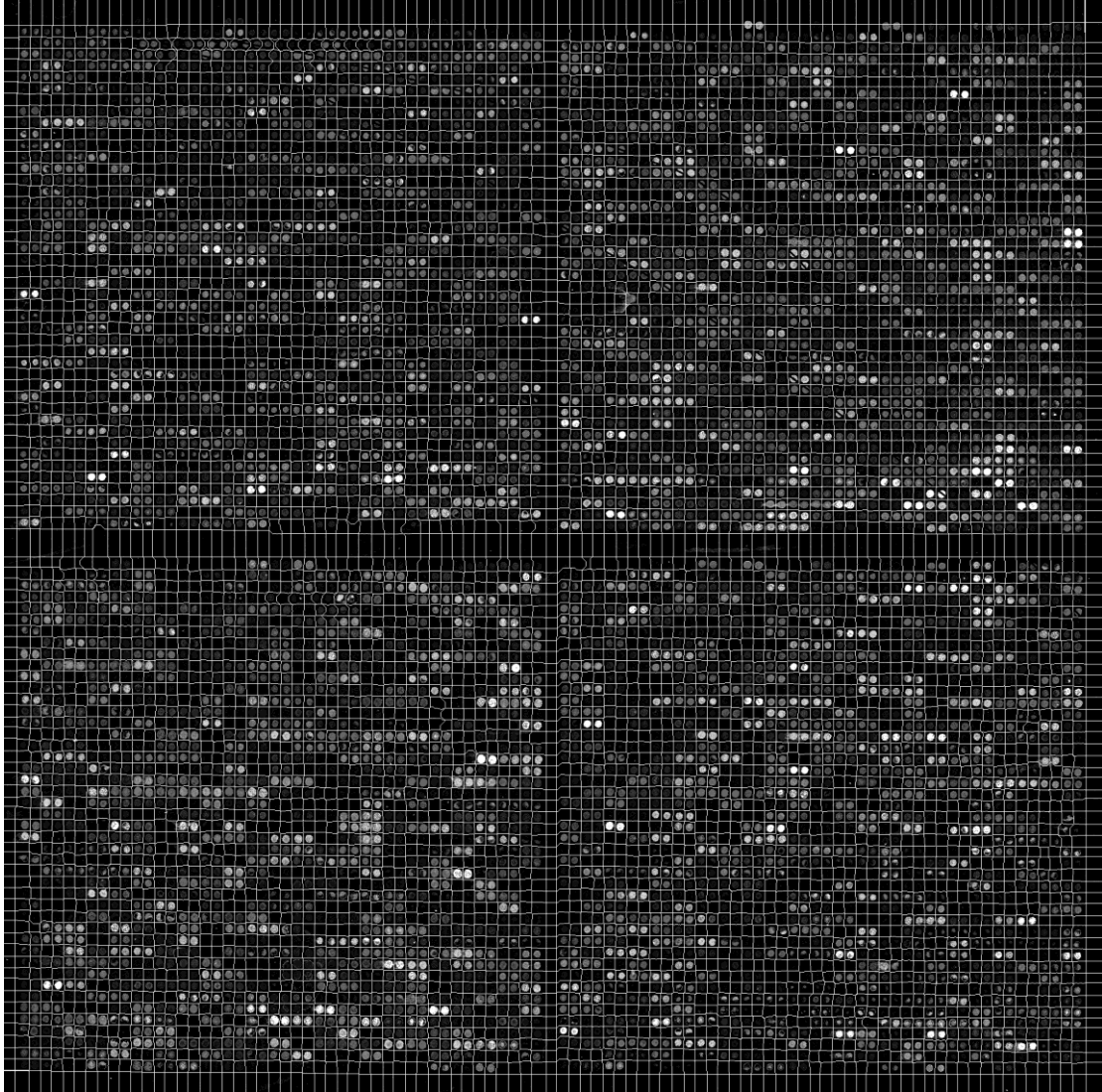


Σχήμα 5.7.β.: Αποτέλεσμα εφαρμογής πλέγματος στην υποεικόνα (2,2)



Σχήμα 5.7.γ.: Αποτέλεσμα εφαρμογής πλέγματος στην υποεικόνα (2,2) σε μεγέθυνση

Προς απόδειξη του ισχυρισμού ότι η προτεινόμενη αυτόματη μεθοδολογία μπορεί να εφαρμόσει πλέγμα σε ολόκληρη εικόνα DNA, η οποία περιέχει υποεικόνες ενώ παράλληλα απαιτείται και η ανίχνευση των υποεικόνων αυτών, παρουσιάζεται στο Σχήμα 5.9, με τέτοια περίπτωση εικόνας με 4 υποεικόνες.



Σχήμα 5.9.: Αποτέλεσμα εφαρμογής πλέγματος σε πραγματική εικόνα DNA με 4 υποεικόνες

Εκμεταλλεούμενος το γεγονός ότι μεταξύ των υποεικόνων, οι γραμμές του πλέγματος απέχουν περισσότερο από τις υπόλοιπες γραμμές του, μπορεί κανείς εύκολα να ανιχνεύσει τις υποεικόνες.

5.3. Συμπεράσματα για τη Προτεινόμενη Μεθοδολογία Αυτόματης Εφαρμογής Πλέγματος

Με απλή εποπτεία των εικόνων που παρουσιάστηκαν στο πέμπτο κεφάλαιο είναι εμφανής η ποιότητα των αποτελεσμάτων της προτεινόμενης μεθοδολογίας. Εκτός από την ποιοτική αξιολόγηση, είναι αναγκαία και μια πιο αντικειμενική, ποσοτική, αξιολόγηση των αποτελεσμάτων, για να είναι εφικτή και η σύγκριση της προτεινόμενης μεθοδολογίας με άλλες υπάρχουσες μεθοδολογίες.

Στον Πίνακα 5.1. παρουσιάζονται κάποια ποσοτικά αποτελέσματα αξιολόγησης της προτεινόμενης αυτόματης μεθοδολογίας για τις εικόνες που παρουσιάστηκαν στο προηγούμενο υποκεφάλαιο, τόσο για τις ιδανικές εικόνες DNA με παραμορφώσεις από αφφινικό και ημιτονοειδή μετασχηματισμό όσο και για τις υποεικόνες από πραγματική εικόνα DNA.

Εικόνα	NS	NS_R	NS_M	Ευστοχία	NP_O	Σφάλμα
Ιδανική εικόνα	400	400	0	100%	0	0%
Αφφινικός-1	400	400	0	100%	0	0%
Αφφινικός-2	400	400	0	100%	0	0%
Ημιτονοειδής-1	400	400	0	100%	0	0%
Ημιτονοειδής-2	400	397	3	99.3%	32	0.01%
Υποεικόνα (1,2)	2041	2025	16	99.2%	1380	0.02%
Υποεικόνα (2,2)	2032	1999	33	98.4%	1490	0.03%
Μέσος Όρος				99.6%		0.01%

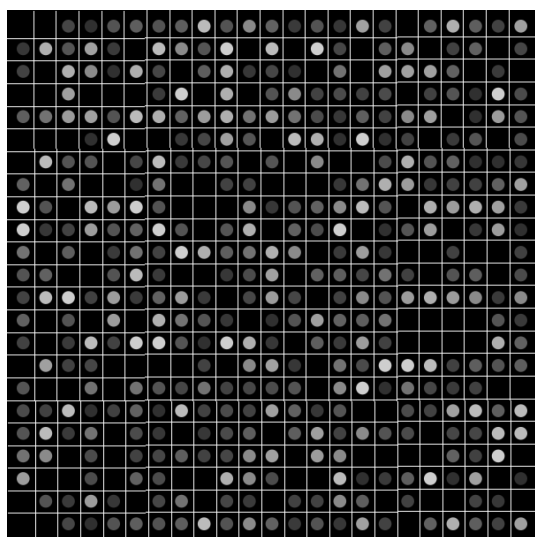
Πίνακας 5.1. Αξιολόγηση αποτελεσμάτων της Προτεινόμενης Αυτόματης Μεθοδολογίας

Η κάθε εικόνα-αποτέλεσμα επιθεωρήθηκε μαζί με το αντίστοιχο πλέγμα και μετρήθηκαν με το χέρι τα στίγματα που ανήκουν ολόκληρα σε κελί του πλέγματος καθώς και τα εικονοστοιχεία (pixels) που ανήκουν σε στίγμα αλλά είναι τοποθετημένα εκτός του αντίστοιχου κελιού του πλέγματος. Η πρώτη στήλη του πίνακα του Πίνακα 5.1, αντιστοιχεί στην εικόνα εισόδου στην οποία εφαρμόστηκε το πλέγμα και η δεύτερη στήλη (NS) αντιστοιχεί στα στίγματα που περιλαμβάνονται στην εικόνα. Η τρίτη στήλη (NS_R) παρουσιάζει το πλήθος των σιγμάτων που περικλείονται πλήρως σε κελί του πλέγματος ενώ η τέταρτη στήλη (NS_M) περιλαμβάνει το πλήθος των σιγμάτων που δεν περικλείονται πλήρως σε κελί του πλέγματος ή που είναι σφαλμένα τοποθετημένα. Η ευστοχία του αλγορίθμου που υπολογίζεται από το πλήθος των σιγμάτων που περικλείονται πλήρως από το πλέγμα (NS_R) ως προς το συνολικό πλήθος των σιγμάτων που περιλαμβάνονται στην εικόνα (NS) φαίνεται στη πέμπτη στήλη. Στην έκτη στήλη (NP_O)

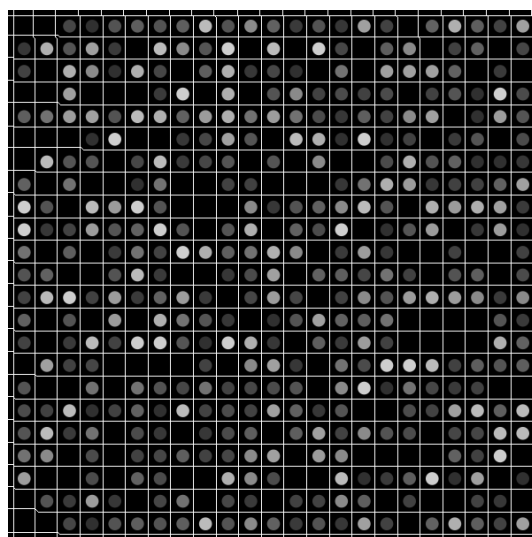
παρουσιάζεται το πλήθος των εικονοστοιχείων (pixels) τα οποία είναι εσφαλμένα τοποθετημένα ή δεν περικλείονται σε κελί του πλέγματος, ενώ η τελευταία στήλη παρουσιάζει το σφάλμα το οποίο υπολογίζεται από το πλήθος των εικονοστοιχείων (pixels) τα οποία είναι εσφαλμένα τοποθετημένα ή δεν περικλείονται σε κελί του πλέγματος (NP_O) ως προς το συνολικό πλήθος των εικονοστοιχείων (pixels) της εικόνας DNA.

5.4. Σύγκριση Προτεινόμενης Αυτόματης Μεθοδολογίας με Μέθοδο Προβολών

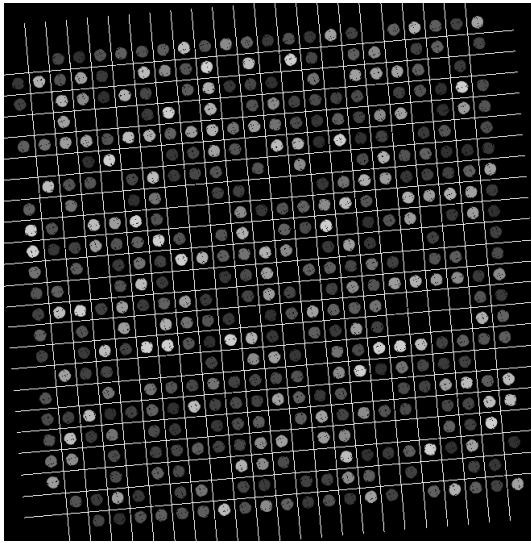
Στην ενότητα αυτή θα παρουσιάζονται το ένα δίπλα από το άλλο, τα αποτελέσματα της Μεθόδου των Προβολών και της Προτεινόμενης Αυτόματης Μεθοδολογίας, για τις ίδιες εικόνες που παρουσιάστηκαν στην προηγούμενη ενότητα, ώστε να είναι εύκολη η σύγκριση τους.



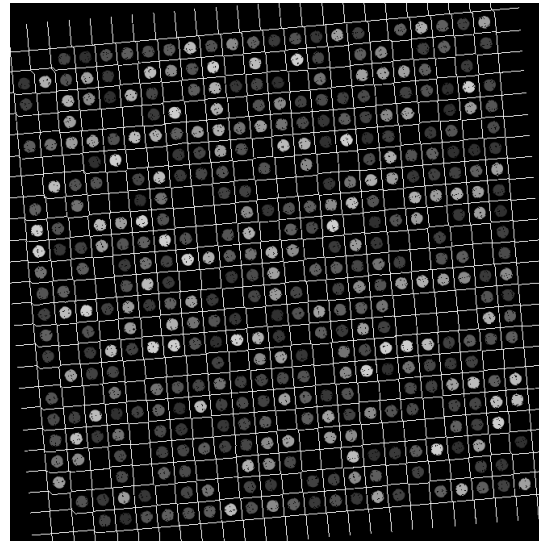
Σχήμα 5.10.α.: Αποτέλεσμα Μεθόδου Προβολών σε ιδανική εικόνα DNA



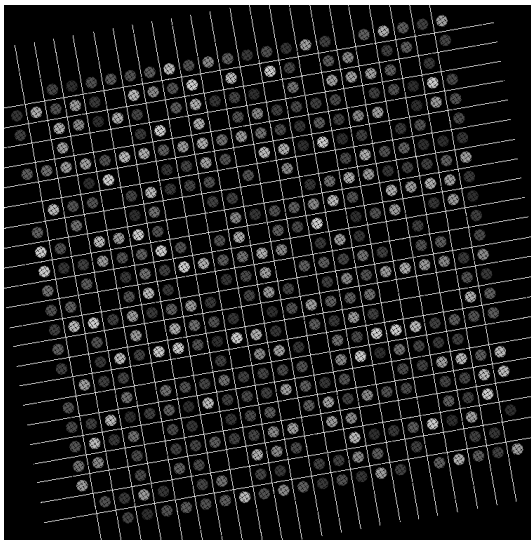
Σχήμα 5.10.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας σε ιδανική εικόνα DNA



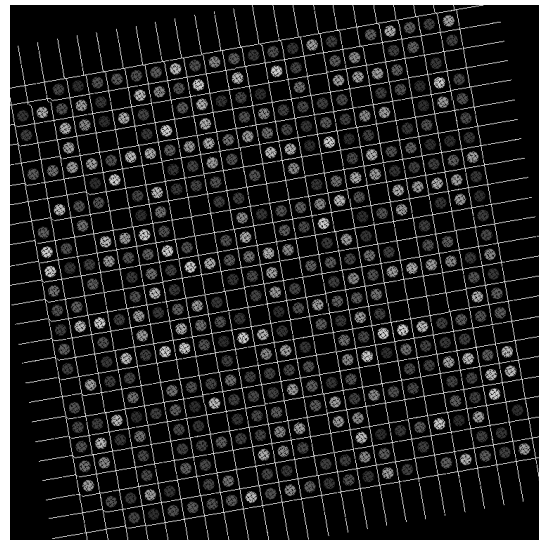
Σχήμα 5.11.α.: Αποτέλεσμα Μεθόδου Προβολών στην εικόνα με Αφρινικό-1



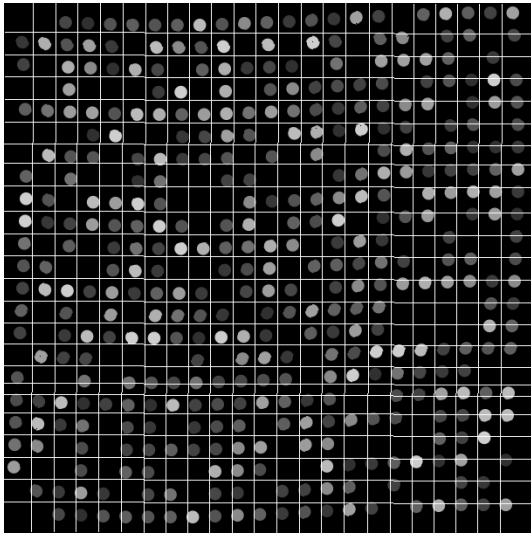
Σχήμα 5.11.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην εικόνα με Αφρινικό-1



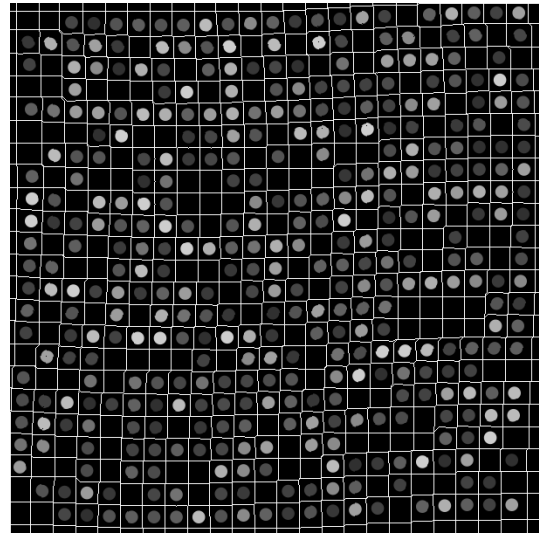
Σχήμα 5.12.α.: Αποτέλεσμα Μεθόδου Προβολών στην εικόνα με Αφρινικό-2



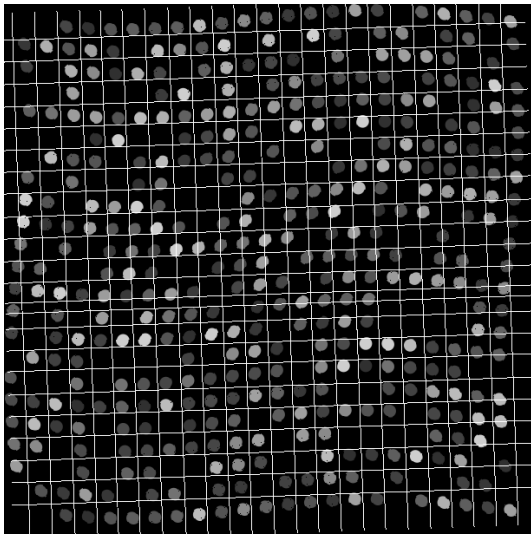
Σχήμα 5.12.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην εικόνα με Αφρινικό-2



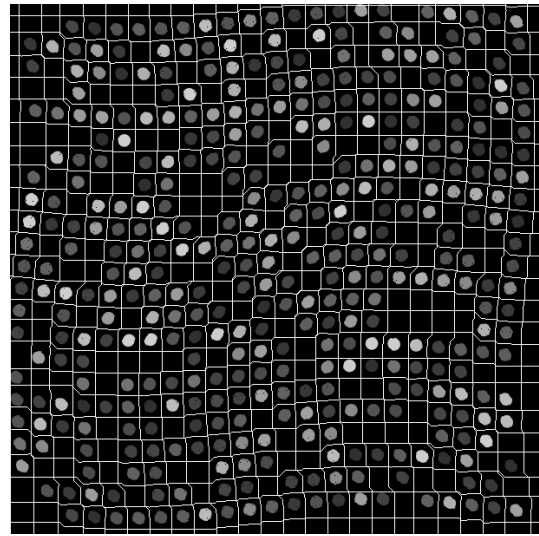
Σχήμα 5.13.α.: Αποτέλεσμα Μεθόδου Προβολών στην εικόνα με Ημιτονοειδή-1



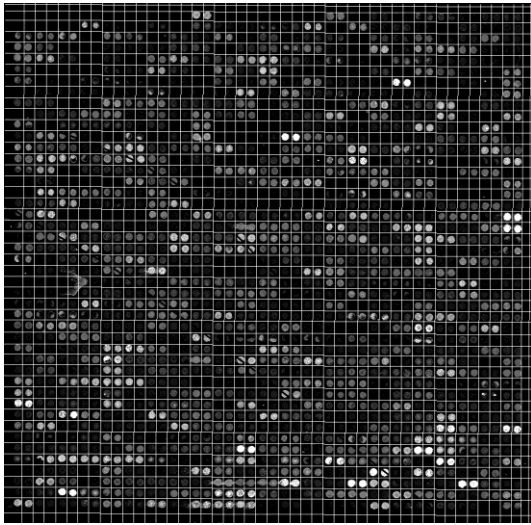
Σχήμα 5.13.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην εικόνα με Ημιτονοειδή-1



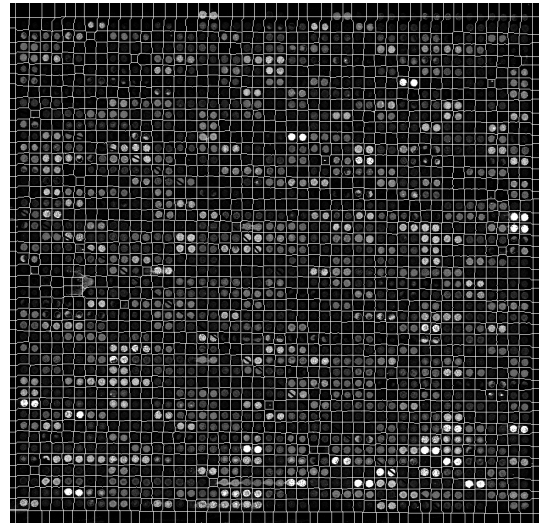
Σχήμα 5.14.α.: Αποτέλεσμα Μεθόδου Προβολών στην εικόνα με Ημιτονοειδή-2



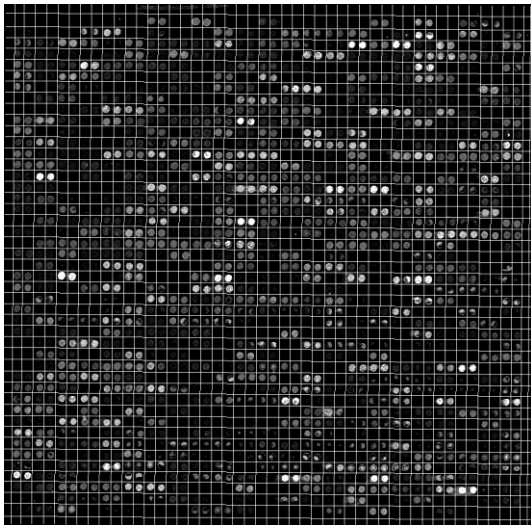
Σχήμα 5.14.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην εικόνα με Ημιτονοειδή-2



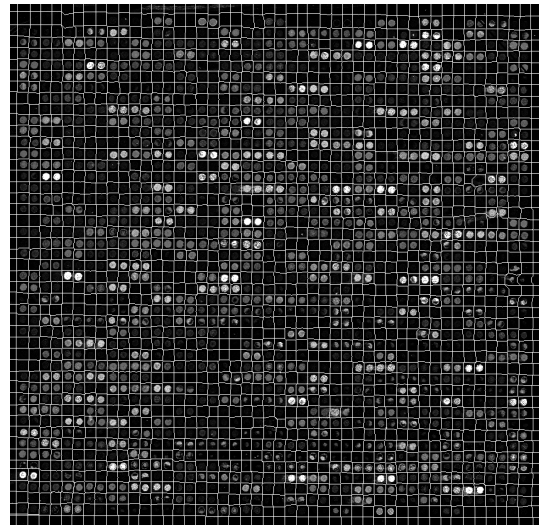
Σχήμα 5.15.α.: Αποτέλεσμα Μεθόδου Προβολών στην υποεικόνα (1,2)



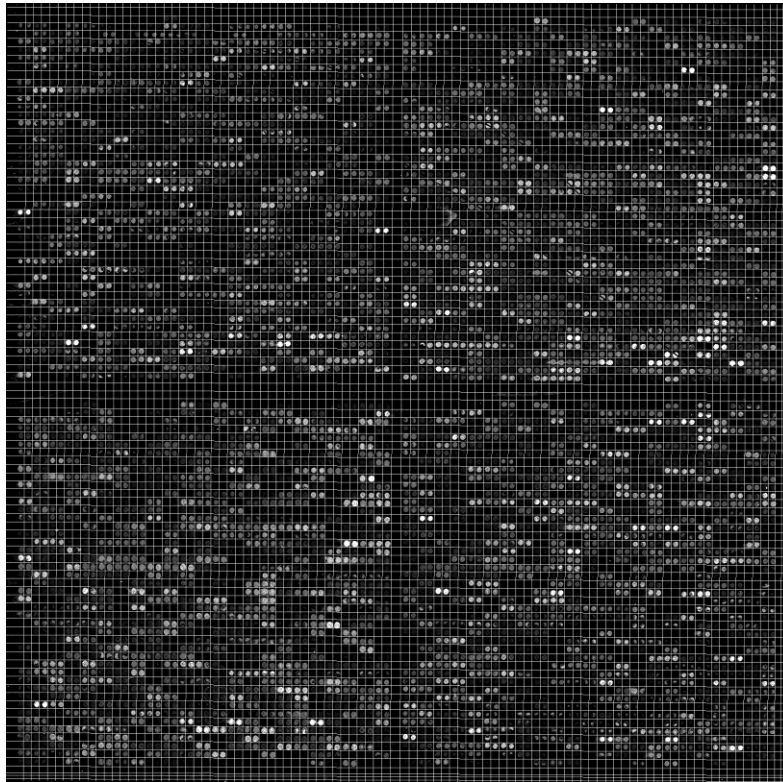
Σχήμα 5.15.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην υποεικόνα (1,2)



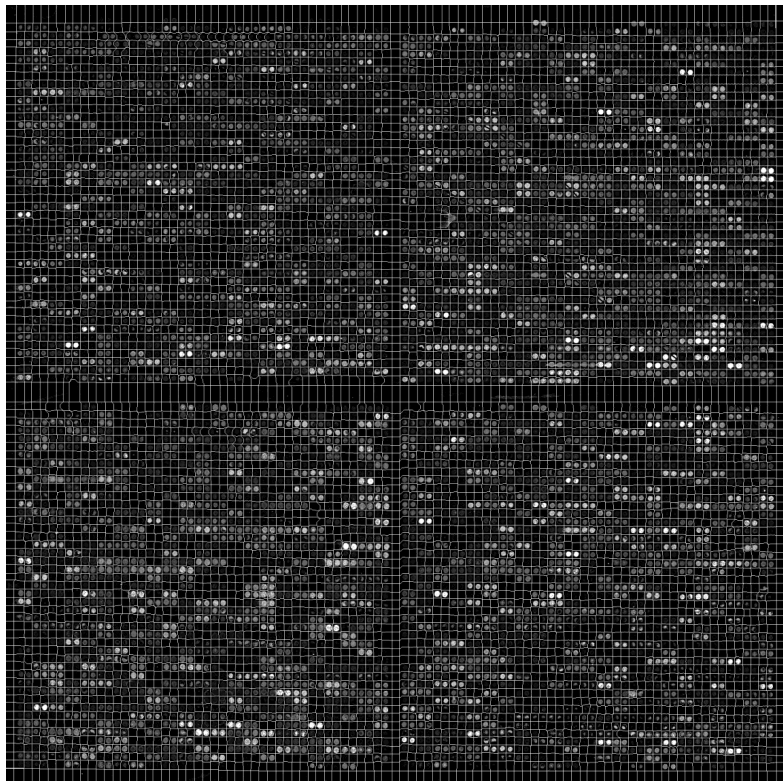
Σχήμα 5.16.α.: Αποτέλεσμα Μεθόδου Προβολών στην υποεικόνα (2,2)



Σχήμα 5.16.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην υποεικόνα (2,2)



Σχήμα 5.17.α.: Αποτέλεσμα Μεθόδου Προβολών στην πραγματική εικόνα DNA με υποεικόνες



Σχήμα 5.17.β.: Αποτέλεσμα Προτεινόμενης Μεθοδολογίας στην πραγματική εικόνα DNA

Με απλή εποπτεία των αποτελεσμάτων των δύο μεθόδων, είναι εμφανής η υπεροχή της προτεινόμενης αυτόματης μεθοδολογίας ως προς την μέθοδο των προβολών, ειδικά στις περιπτώσεις εικόνων που παραμορφώθηκαν από ημιτονοειδή μετασχηματισμό, κατά τις οποίες η δεύτερη δεν μπορεί να αντεπεξέλθει.

Για να είναι όμως εφικτή η ποσοτική σύγκριση των αποτελεσμάτων των δύο μεθόδων, είναι αναγκαίο να συμπληρωθεί και για την μέθοδο των προβολών, ένας πίνακας με ποσοτικά αποτελέσματα, όπως τον Πίνακα 5.1. που παρουσιάστηκε για την προτεινόμενη μεθοδολογία στην ενότητα 5.3 (βλ. Πίνακας 5.2).

Εικόνα	NS	NS_R	NS_M	Ευστοχία	NP_O	Σφάλμα
Ιδανική εικόνα	400	327	73	81.8%	5840	1.6%
Αφφινικός-1	400	327	73	81.8%	5840	1.6%
Αφφινικός-2	400	327	73	81.8%	5840	1.6%
Ημιτονοειδής-1	400	212	188	53.0%	16920	4.7%
Ημιτονοειδής-2	400	162	238	40.5%	28200	8,7%
Υποεικόνα (1,2)	2041	1925	116	94.3%	5870	0.7%
Υποεικόνα (2,2)	2032	1883	249	92.6%	7257	0.9%
Μέσος Όρος				75.1%		2.83%

Πίνακας 5.2. Αξιολόγηση αποτελεσμάτων της Μεθόδου των Προβολών

Ο Πίνακας 5.3, περιλαμβάνει τα συγκριτικά ποσοτικά αποτελέσματα για τις δύο μεθόδους.

	Ευστοχία		Σφάλμα	
	ΜΠ	ΠΑΜ	ΜΠ	ΠΑΜ
Ιδανική εικόνα	81.8%	100%	1.6%	0%
Αφφινικός-1	81.8%	100%	1.6%	0%
Αφφινικός-2	81.8%	100%	1.6%	0%
Ημιτονοειδής-1	53.0%	100%	4.7%	0%
Ημιτονοειδής-2	40.5%	99.3%	8,7%	0.01%
Υποεικόνα (1,2)	94.3%	99.2%	0.7%	0.02%
Υποεικόνα (2,2)	92.6%	98.4%	0.9%	0.03%
Μέσος Όρος	75.1%	99.6%	2.83%	0.01%

Πίνακας 5.3. Συγκριτικός Πίνακας Αποτελεσμάτων

Στον Πίνακα 5.3, η συντομογραφία **ΜΠ** αντιπροσωπεύει την Μέθοδο Προβολών ενώ **ΠΑΜ** είναι η Προτεινόμενη Αυτόματη Μεθοδολογία. Η υπεροχή της προτεινόμενης αυτόματης μεθοδολογίας εφαρμογής πλέγματος, είναι εμφανής ακόμη και στα ποσοτικά αποτελέσματα.

Κεφάλαιο

6

Αξιολόγηση

6.1. Αξιολόγηση της Προτεινόμενης Αυτόματης Μεθοδολογίας

Σε κάθε περίπτωση υλοποίησης ενός αλγορίθμου, ο προγραμματιστής αναγκάζεται να επιλέξει (“tradeoff”) μεταξύ ταχύτητας και ποιότητας αποτελεσμάτων. Όσο κυνηγά κανείς περισσότερη ποιότητα, θυσιάζει ταχύτητα εκτέλεσης όπως και αντίστροφα, όσο προσπαθεί κανείς να κάνει πιο γρήγορο τον αλγόριθμο, θυσιάζει από την ποιότητα των αποτελεσμάτων.

Στην προκειμένη περίπτωση, επιλέχθηκε η κατεύθυνση της καλύτερης ποιότητας στα αποτελέσματα. Αυτό βέβαια κόστισε στη ταχύτητα εκτέλεσης των αποτελεσμάτων. Ενώ οι ήδη υπάρχουσες μεθοδολογίες πλησιάζουν χρόνους εκτέλεσης τάξης μεγέθους δευτερολέπτων, η προτεινόμενη μέθοδος φτάνει τάξη μεγέθους λεπτών. Αν και μεγάλη η διαφορά, επειδή η μέθοδος αυτή εκτελείται μια φορά για κάθε εικόνα DNA, αξίζει να θυσιάσει κανείς ταχύτητα για να πάρει το βέλτιστο αποτέλεσμα.

6.2. Μελλοντικές Βελτιώσεις και Επεκτάσεις

Ένα καίριο σημείο της προτεινόμενης μεθοδολογίας είναι η ανίχνευση των κέντρων των στιγμάτων η οποία όπως περιγράφηκε στην παράγραφο 4.1.4., επιτυγχάνεται με ένα αλγόριθμο ταιριάσματος προτύπων στην εικόνα DNA.

Οι πραγματικές εικόνες DNA υποφέρουν από θόρυβο στο υπόβαθρο (background). Όσο επιτυχής και να είναι όμως το φιλτράρισμα, ο θόρυβος προκαλεί πολλά προβλήματα στον αλγόριθμο ταιριάσματος προτύπων. Για να αποφευχθούν τέτοιου είδους προβλήματα, επιλέγονται τα κέντρα στιγμάτων των οποίων οι συντεταγμένες ανήκουν στη λευκή περιοχή της εικόνας με τις υποεικόνες. Με τον τρόπο αυτό όμως υπάρχει πιθανότητα να μην επιλεγούν όλα τα σωστά κέντρα που ανιχνεύτηκαν.

Ένας πιο αντικειμενικά σωστός τρόπος, είναι να εκμεταλλευτεί κανείς την διαφορετική “υφή” της περιοχής των σωστών κέντρων σε αντίθεση με την περιοχή των εσφαλμένων κέντρων στιγμάτων. Η χωρική κατανομή των εσφαλμένων κέντρων διαφέρει κατά πολύ από την, σταθερής απόστασης, χωρική κατανομή των σωστών κέντρων στιγμάτων, τα οποία σε κατάλληλες συνθήκες είναι ζυγισμένα – στοιχισμένα. Η στατιστική είναι ένα πολύ δυνατό εργαλείο για επίλυση τέτοιων προβλημάτων και η χρήση εργαλείων της, όπως για παράδειγμα “statistical outliers”, θα βελτιώσουν κατά πολύ την όλη μεθοδολογία.

Βιβλιογραφία

- [1] *Dilation (morphology)*. (n.d.). Ανάκτηση June 20, 2011, από Wikipedia: [http://en.wikipedia.org/wiki/Dilation_\(morphology\)](http://en.wikipedia.org/wiki/Dilation_(morphology))
- [2] *DNA microarrays technology*. (n.d.). Ανάκτηση June 10, 2011, από Bioinformatics Tutorials: <http://www.bioinformaticstutorials.com>
- [3] *DNA microarrays technology*. (n.d.). Ανάκτηση June 10, 2011, από National Human Genome Research Institute: <http://www.genome.gov>
- [4] Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification, 2nd edition*. New York: Wiley-Interscience.
- [5] *Image Spatial Data Analysis Group*. (n.d.). Ανάκτηση June 11, 2011, από University of Illinois at Urbana-Champaign: <http://isda.ncsa.uiuc.edu/Microarrays/>
- [6] Karsten, H., & Jens, M. C. (February 2003). Bayesian Grid Matching. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE* , Vol.25 No.2.
- [7] Luis, R., & Vidya, V. (January-March 2006). A Hill-Climbing Approach for Automatic Gridding of cDNA Microarray Images. *IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS* , Vol.3 No.1.
- [8] *Minimum Information About a Microarray Experiment - MIAME*. (n.d.). Ανάκτηση June 12, 2011, από Microarray Gene Expression Data (MGED) Society: <http://www.mged.org/>
- [9] Norbert, B., Hilmar, L., & Horst, B. (n.d.). Automatic Grid Fitting for Genetic Spot Array Images Containing Guide Spots. *Novartis Forschungsinstitut Genetics* .
- [10] Oliveros, A., & Sotaquira, M. (July 2008). An Automatic Gridding and Contour Based Segmentation Approach Applied to DNA Microarray Image Analysis. *PROCEEDINGS OF WORLD ACADEMY OF SCIENCE* , Vol.31.
- [11] Rafael, C. G. (2011). *ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΑΣ, 3rd edition*. ΑΘΗΝΑ: ΤΖΙΟΛΑ.
- [12] Yuk, F. L., & Duccio, C. (November 2003). Fundamentals of cDNA microarray data analysis. *TRENDS in Genetics* , Vol.19 No.11.
- [13] Γιόβλα, Δ. (2004). *ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΒΙΟΪΑΤΡΙΚΗ ΚΑΙ ΚΥΤΤΑΡΙΚΗ ΜΗΧΑΝΙΚΗ*. ΑΘΗΝΑ: ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ.