



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
"ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ"**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Επεξεργασία εικόνων μικροσυστοιχιών cDNA με εύρωστες
τεχνικές αυτόματης ταξινόμησης**

Μαργαρίτα Γ. Μαγιάτη

ΕΠΙΒΛΕΠΩΝ: Διονύσιος Κάβουρας, Καθηγητής

ΑΘΗΝΑ

ΑΠΡΙΛΙΟΣ 2012

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Επεξεργασία εικόνων μικροσυστοιχιών cDNA με εύρωστες τεχνικές αυτόματης ταξινόμησης

Μαργαρίτα Γ. Μαγιάτη

A.M.: ΠΙΒ036

ΕΠΙΒΛΕΠΩΝ: Διονύσιος Κάβουρας, Καθηγητής

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: Ερρίκος Βεντούρας, Καθηγητής
Διονύσιος Κάβουρας, Καθηγητής
Εμμανουήλ Σαγκριώτης, Αναπληρωτής Καθηγητής

Απρίλιος 2012

ΠΕΡΙΛΗΨΗ

Σκοπός της παρούσας εργασίας ήταν, σε πρώτο στάδιο, η διερεύνηση της αποτελεσματικότητας διάφορων φίλτρων κυματίων, για την πιο αποτελεσματική καταστολή του θορύβου σε εικόνες μικροσυστοιχιών cDNA. Σε δεύτερο στάδιο ο σκοπός ήταν η υλοποίηση ενός αλγορίθμου κατάτμησης για την πιο αποτελεσματική ανάδειξη των κηλίδων από τον περιβάλλοντα φόντο. Η καινοτομία της παρούσας διπλωματικής έγκειται στο γεγονός ότι ο προτεινόμενος τροποποιημένος αλγόριθμος MRF χρησιμοποιεί ένα επιπλέον χαρακτηριστικό, που προκύπτει μετά την εφαρμογή του σταθμευμένου μετασχηματισμού κυματίων SWT στις εικόνες.

Στην παρούσα διπλωματική εργασία χρησιμοποιήθηκαν εικόνες που προέκυψαν από πειράματα μικροσυστοιχιών cDNA, ως πρότυπα για τη δημιουργία προσομοιωμένων εικόνων σε πέντε διαφορετικά επίπεδα θορύβου, για την αξιολόγηση των αλγορίθμων κατάτμησης που υλοποιήθηκαν. Επιπλέον, πραγματοποιήθηκε μία συστηματική εκτίμηση φίλτρων βασισμένα σε κυμάτια για την καταστολή του θορύβου, προκειμένου να επιτευχθεί βελτίωση των εικόνων μικροσυστοιχιών cDNA. Για το σκοπό αυτό έγινε χρήση του SWT με διάφορα είδη μητρικών κυματίων. Οι εικόνες που προέκυψαν μετά την καταστολή του θορύβου αναλύθηκαν και εφαρμόστηκε σε αυτές ο αλγόριθμος κατάτμησης τυχαίων πεδίων Μαρκόφ MRF, για την εκτίμηση του αντίκτυπου της διαδικασίας καταστολής θορύβου στο στάδιο της κατάτμησης. Σκοπός αυτής της διαδικασίας ήταν η επιλογή των κατάλληλων παραμέτρων, κυρίως όσον αφορά το είδος του κυματίου, ούτως ώστε να χρησιμοποιηθούν στην μετέπειτα ανάλυση. Επιπλέον υλοποιήθηκε ένας ημιαυτόματος αλγόριθμος δημιουργίας και ευθυγράμμισης πλέγματος. Όσον αφορά την κατάτμηση των εικόνων, υλοποιήθηκαν τρεις αλγόριθμοι: ο k-means, ο MRF και μία τροποποίηση του MRF με χρήση των κυματίων (Wavelet-MRF). Για ποσοτικοποίηση των αποτελεσμάτων, όσον αφορά την καταστολή του θορύβου μετρήθηκε το mean square error (MSE) και το signal/MSE. Όσον αφορά την απόδοση των τεχνικών κατάτμησης, υπολογίστηκαν ο παράγοντας ταυτοποίησης κατάτμησης SMF και ο συντελεστής προσδιορισμού r^2 .

Σύμφωνα με τα αποτελέσματα, πιο αποτελεσματική καταστολή του θορύβου κατάφερε να επιτύχει η τεχνική ελαφρού φιλτραρίσματος σε συνδυασμό με το biorthogonal 1.3 μητρικό κυμάτιο, επιτυγχάνοντας το χαμηλότερο MSE και τα υψηλότερα signal/MSE και SMF. Όσον αφορά την κατάτμηση των εικόνων ο k-means επιτυγχάνει το χαμηλότερο SMF και το χαμηλότερο r^2 , ακολουθεί ο MRF και την καλύτερη απόδοση παρουσιάζει ο Wavelet-MRF. Σε υψηλά επίπεδα θορύβου οι αλγόριθμοι MRF και Wavelet-MRF παρουσιάζουν μεγάλη διαφορά σε σχέση με τον k-means, ενώ σε χαμηλά επίπεδα θορύβου η διαφορά είναι αρκετά μικρότερη, αποδεικνύοντας ότι οι αλγόριθμοι αυτοί ανταποκρίνονται καλύτερα σε υψηλά επίπεδα θορύβου. Επιπλέον, η χρήση των κυματίων στον αλγόριθμο Wavelet-MRF βελτιώνει το αποτέλεσμα του MRF, επιβεβαιώνοντας τη σημαντικότητα της πληροφορίας που παρέχεται από τα κυμάτια, καθώς και την αναγκαιότητα για διερεύνηση περισσότερων χαρακτηριστικών που θα μπορούσαν να συνεισφέρουν αποτελεσματικά στον αλγόριθμο κατάτμησης MRF για περαιτέρω βελτίωση της απόδοσής του.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Βιοπληροφορική, Επεξεργασία Εικόνας

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: μικροσυστοιχίες cDNA, κυμάτια, κατάτμηση, τυχαία πεδία Μαρκόφ, μετασχηματισμός κυματίων

ABSTRACT

The purpose of the present study was initially to investigate the effectiveness of different wavelet filters, for the most effective suppression of noise in microarray cDNA images. In the second stage the objective was the implementation of a segmentation algorithm for a more effective delineation of the spots from the surrounding background. The innovation of this thesis is that the proposed modified Markov random fields' algorithm (MRF) uses an additional feature, obtained by the application of the stationary wavelet transform (SWT) to images.

In this thesis, images derived from cDNA microarray experiments were used as templates for creating simulated images at five different levels of noise, for the evaluation of the implemented segmentation algorithms. Moreover, a systematic evaluation of filters based on wavelets was performed, in order to achieve the improvement of microarray cDNA images by means of noise suppression. For this purpose, SWT was used combined with various types of mother wavelets. The images obtained after the suppression of noise were analyzed and the MRF algorithm was applied to them in order to assess the impact of the noise suppression process to the stage of segmentation. The purpose of this procedure was the selection of the appropriate parameters, basically concerning the type of wavelet, in order to use them in later analysis. Moreover, a semiautomatic algorithm for the gridding of the images was implemented. Regarding the segmentation of images, three algorithms were implemented: the k-means, the MRF and a modification of the MRF using wavelets (Wavelet-MRF). For the quantification of the results in terms of noise suppression the mean square error (MSE), and signal/MSE were measured. Regarding the performance of segmentation techniques, the segmentation matching factor SMF and the coefficient of determination r^2 were calculated.

According to the results, more effective noise suppression was achieved by the soft thresholding technique combined with the biorthogonal 1.3 mother wavelet, scoring the lowest MSE and the highest signal/MSE and SMF. Regarding the segmentation of images, the k-means algorithm achieved the lowest SMF and the lowest r^2 , followed by the MRF, while the Wavelet-MRF algorithm showed the best performance. At high levels of noise, MRF and Wavelet-MRF algorithms show significant difference compared to the k-means, while at low noise levels the difference is much smaller, showing that these algorithms are better suited to high noise levels. Furthermore, the use of wavelets in the Wavelet-MRF algorithm improves the results of the MRF, confirming the importance of the information provided by wavelets, as well as the need to explore more features that could contribute effectively to the MRF segmentation algorithm, for further improvement of its segmentation accuracy.

SUBJECT AREA: Bioinformatics, Image Processing

KEYWORDS: cDNA microarrays, wavelets, segmentation, Markov random fields, stationary wavelet transform