

**Τίτλος:**  
**Ανάλυση δομικών τανυστών σε πρωτεΐνες: αποτελεσματική εξαγωγή  
χαρακτηριστικών για πρόβλεψη ετεροπολυμερών συστοιχιών**

**Περίληψη:**

Οι τρέχουσες τεχνικές ανάλυσης της μακρομοριακής δομής αδυνατούν να γενικεύσουν καλά σε διαφορετικούς τύπους δομών, καθώς κάθε μία από αυτές τις μεθόδους είναι σχεδιασμένη για συγκεκριμένους τύπους δομικών στοιχείων. Το Integrative Modeling είναι ένας τρόπος για να συνδυάσουμε τα πειραματικά με τα υπολογιστικά δεδομένα, αλλά εμφανίζει κάποιες αδυναμίες όσον αφορά στα ετεροπολυμερικά σύμπλοκα (ειδικότερα στα μη-συμμετρικά), όπου οι επιμέρους μονάδες τους μπορεί να υιοθετούν διαφορετικές διαμορφώσεις ανάλογα με το αν είναι απομονωμένες ή μέσα στη συστοιχία τους. Στην εργασία αυτή, εκμεταλλευόμαστε το πεδίο της μηχανικής όρασης, και χρησιμοποιούμε local descriptors και τα βασικά βήματα του αλγορίθμου SIFT για την εξαγωγή διακριτών χαρακτηριστικών (τοπικά ακρότατα) από εικόνες: ανιχνεύουμε τα σημεία-κλειδιά (keypoints) στα density maps των ατομικών δομών, ώστε να τις εντοπίσουμε μέσα στους cryo-EM maps των μακρομοριακών τους συστοιχιών. Για τη μείωση του τεράστιου πλήθους από keypoints, εστιάζουμε στην εύρεση γωνιών, καθώς τα σημεία αυτά παραμένουν σταθερά ανεξαρτήτου περιστροφής ή αλλαγής. Καθώς οι ιδιοτιμές και τα αντίστοιχα ιδιοδιανύσματα της δομής του τανυστή περιγράφουν τη βασική καμπυλότητα της δομής, υιοθετούμε τις αρχές της μεθόδου ανίχνευσης γωνιών Harris, και τις επεκτείνουμε σε 3-D ανάλυση δομικών τανυστών. Βασιζόμενοι στις στατιστικές των λόγων των ιδιοτιμών, εφαρμόζουμε πολλαπλούς τύπους κατωφλίωσης για διαφορετικές παραμέτρους, και δοκιμάζουμε αυτές τις παραμέτρους σε 54 διαφορετικές δομές. Για την αξιολόγηση των παραμέτρων, συγκρίνουμε τα υπολογισθέντα keypoints με ένα σύνολο για το οποίο γνωρίζουμε ότι επιτυγχάνει σωστή πρόβλεψη συστοιχιών. Τα πειραματικά αποτελέσματα δείχνουν την ύπαρξη παραμέτρων που αφαιρούν σχεδόν όλα τα ασταθή keypoints (false positives), παραμέτρων που διατηρούν σχεδόν όλα τα σταθερά (true positives), και παραμέτρων που δίνουν λύσεις εξισορροπώντας το trade-off μεταξύ των προηγούμενων δύο. Τέλος, επαληθεύουμε ότι υπάρχουν σύμπλοκα με αναξιόπιστο προφίλ πυκνότητας, καθώς δε υπάρχουν λύσεις για όλες τις αναλύσεις τους. Η μέθοδος που προτείνουμε είναι ένας γενικός, γρήγορος και ακριβής τρόπος για την εξαγωγή τοπικών χαρακτηριστικών για σωστή πρόβλεψη συστοιχίας, και μπορεί να χρησιμεύσει ως βασική γραμμή για τη μελέτη των δυναμικών αυτών των keypoints όταν υπόκεινται σε διαμορφωτικές αλλαγές.

**Title:**  
**Structure tensor analysis on proteins: efficient feature extraction for heteromultimeric assembly prediction**

**Abstract:**

Current macromolecular structure resolution techniques fail to generalize well across different types of structures, since each method is designed for specific kinds of components. Integrative Modeling combines experimental and computational data regardless of their resolution, but it also has limitations when it comes to heteromultimeric complexes (especially asymmetric), where the individual components may adopt different conformations whether they are isolated or within their assembly. In this work, we exploit the field of computer vision, and we make use of local descriptors and the main steps of SIFT algorithm, for extracting distinctive features (local extrema) from images; we detect informative features (keypoints) in the atomic structures' density maps, so as to localize them within their macromolecular assembly's cryo-EM maps. For diminishing the huge number of extracted features, we specifically search for corners, as these points remain stable regardless rotation or change. Since the eigenvalues and the eigenvectors of the structure tensor describe the principal curvatures of the structure, we adopt the principles of Harris corner detector and extend them to a 3D structure tensor analysis (STA). Based on the statistics of the eigenvalues' ratios, we apply multiple types of thresholding under different configurations, and benchmark the STA set of parameters on 54 different structures. For the parameter evaluation, we compare the extracted keypoints with a set that is known to lead to correct assembly prediction. Experimental results show the existence of parameters that remove almost all of the unstable keypoints (false positives), others that retain almost all of the stable ones (true positives), and others that can balance the trade-off between those two. Finally, we verify that there are specific complexes without a trustworthy density profile, since no solutions can be obtained for every resolution. The proposed method is a general, fast, and accurate way for extracting localized features for correct assembly prediction, which can serve as a baseline for studying the dynamics of these keypoints under conformational changes.