



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Γραφοθεωρητικές Μέθοδοι Όρασης Υπολογιστών
και Εκμάθησης Μηχανών για Ανίχνευση
Αντικειμένων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Χρήστου Γ. Μπάμπη

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2014



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Γραφοθεωρητικές Μέθοδοι Όρασης Υπολογιστών
και Εκμάθησης Μηχανών για Ανίχνευση
Αντικειμένων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Χρήστου Γ. Μπάμπη

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 23η Ιουλίου 2014.

.....
Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

.....
Γεώργιος Παπαβασιλόπουλος
Καθηγητής Ε.Μ.Π.

.....
Γεράσιμος Ποταμιάνος
Αν. Καθηγητής Π.Θ

Αθήνα, Ιούλιος 2014

.....
Χρήστος Γ. Μπάμπης

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

Copyright © Χρήστος Γ. Μπάμπης, 2014

Με την 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 το συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η σύγχρονη εποχή μπορεί με μια φράση να χαρακτηριστεί ως η εποχή της πληροφορίας. Η αποτελεσματική συλλογή, ανάλυση και αποθήκευση αυτής κρίνονται αναγκαίες αν αναλογιστεί κανείς τον όγκο, την πολυπλοκότητα στη δομή και την αναγκαιότητα γρήγορης επεξεργασίας αυτής. Επομένως, η αξιοποίηση νέων μεθόδων αναπαράστασης της πληροφορίας κρίνεται επιτακτική. Η κυρίαρχη δομή η οποία καλύπτει την παραπάνω ανάγκη είναι οι γράφοι.

Στα προβλήματα της Όρασης Υπολογιστών, συναντάει κανείς την πληροφορία υπό μορφή εικόνων. Τα τελευταία χρόνια, έχει δοθεί μια όλο και αυξανόμενη βαρύτητα στην εφαρμογή γραφοθεωρητικών μεθόδων για την επεξεργασία και ανάλυση των εικόνων με απώτερο σκοπό την κατανόηση του περιεχομένου της εικόνας. Η προσαρμοστικότητα της αφαιρετικής δομής των γράφων έδωσε, επομένως, βήμα σε νέες τεχνικές και μεθόδους.

Στόχος της διπλωματικής εργασίας είναι η μελέτη και διερεύνηση αυτών των νέων ερευνητικών κατευθύνσεων υπό το πρίσμα της χρήσης όχι του κλασσικού κανονικού πλέγματος για την αναπαράσταση μιας εικόνας, αλλά μέσω ενός γράφου στον οποίο οι κόμβοι είναι αυθαίρετα τοποθετημένοι. Παρουσιάζονται βασικά γραφοθεωρητικά στοιχεία καθώς και γνωστές τεχνικές από τη βιβλιογραφία οι οποίες στοχεύουν σε μια από τις πιο βασικές διεργασίες της Όρασης Υπολογιστών, την κατάτμηση αντικειμένων. Οι δύο βασικές κατηγορίες των τεχνικών που περιγράφονται είναι οι μέθοδοι με επίβλεψη και χωρίς επίβλεψη.

Η συνδρομή της εργασίας στις κατευθύνσεις αυτές είναι διπλή. Στο κομμάτι των μεθόδων με επίβλεψη, ενοποιείται η διάχυση θερμότητας πάνω σε γράφους με το επιδημιολογικό μοντέλο SIR. Με βάση αυτή τη σύνδεση, προτείνεται μια βελτίωση του αλγόριθμου του Τυχαίου Περιπατητή η οποία μπορεί να δώσει καλύτερα αποτελέσματα στις περισσότερες των περιπτώσεων. Αναφορικά με τις μεθόδους δίχως επίβλεψη, προτείνεται η εφαρμογή της παραγοντοποίησης Cluster Non-Negative Matrix Factorization (Cluster NMF) με σκοπό την κατάτμηση εικόνας. Τέλος, επεκτείνεται το σύστημα αυτό με τη βοήθεια ιδεών από Consensus Learning. Η χρήση των μεθόδων αυτών στις εικόνες ήταν μέχρι πρότενος απαγορευτική λόγω της μεγάλης διάστασης του χώρου των χαρακτηριστικών. Η χρήση ενός γράφου κατάλληλα επιλεγμένου και η αξιοποίησή του, επιτρέπουν πλέον την εφαρμογή τέτοιων μεθόδων.

Abstract

The modern age is usually characterized as the information age. The efficient collection, analysis and storage of information are necessary considering the volume, complexity, structure of information and the need to perform these processes in a fast way. Therefore, the use of new methods of representation of information is imperative. A very promising structure that can meet the above requirements are graphs.

In Computer Vision problems, images are the main source of information. In recent years, there has been an increasing emphasis on the application of graph theoretic methods for the processing and analysis of images in order to understand the contents of an image. As a result, the versatile graph structure has given rise to new methods and techniques.

The aim of this thesis is to study and explore these new research directions by using a non regular grid to represent an image. An introduction to graph theoretic methods and state of the art techniques applied for object detection, a very common task in Computer Vision, are presented. The two main categories of techniques described are supervised methods and unsupervised ones.

The contribution of this thesis with respect to these directions is twofold. Concerning supervised methods, the epidemiological model SIR is theoretically proven to result in the heat diffusion schemes on graphs. Based on this connection, an improvement of the Random Walker algorithm is suggested, which can give better results in most cases. Regarding unsupervised methods, the Cluster Non-Negative Matrix Factorization (Cluster NMF) method is proposed as a new tool for the purpose of image segmentation. Finally, a new segmentation system based on the Consensus Learning framework is studied and implemented. These methods were until now computationally intractable due to the large dimension of the feature space involved. The use of an appropriately defined graph structure now allows the application of such methods.

Ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω τον κ. Πέτρο Μαραγκό για την επίβλεψη της διπλωματικής μου εργασίας. Με τις ιδέες και την εμπειρία τις οποίες μου προσέφερε στα επιμέρους στάδια της εργασίας, συνέβαλλε στην εξερεύνηση αλλά και την ουσιαστική μελέτη νέων ερευνητικών κατευθύνσεων. Τέλος, θα ήθελα να ευχαριστήσω θερμά την οικογένεια και τους φίλους μου, οι οποίοι με τη στήριξη και το ενδιαφέρον τους συνέδραμαν με ουσιαστικό τρόπο στην ολοκλήρωση της εργασίας.

Περιεχόμενα

1	Εισαγωγή	13
1.1	Η πληροφορία στη σύγχρονη εποχή	13
1.2	Όραση Υπολογιστών και γραφοθεωρητικές μέθοδοι	14
1.3	Κεντρική ερευνητική κατεύθυνση και συμβολές της διπλωματικής εργασίας	14
2	Κατασκευή Γράφων και Ρύθμιση Παραμέτρων	15
2.1	Εισαγωγή στους γράφους	15
2.2	Κατασκευή γράφου για εφαρμογές ανίχνευσης αντικειμένων	17
2.2.1	Τυπικές συναρτήσεις ομοιότητας	17
2.2.2	Μέθοδος Linear Neighborhood Propagation (LNP)	18
2.2.3	Μέθοδος Global Linear Neighborhood Propagation (GLNP)	20
2.3	Αυτοματοποιημένη ρύθμιση παραμέτρων	22
2.3.1	Περιγραφή της βασικής ιδέας	22
2.3.2	Επέκταση σε 3 διάστασεις - Σχόλια	25
2.4	Συμπεράσματα κεφαλαίου 2 - Ανακεφαλαίωση	27
3	Ελαχιστοποίηση Συναρτησιακών επί του Γράφου - Diffusion	28
3.1	Μοντέλο επιδημιολογικής εξάπλωσης SIR	28
3.1.1	Θεωρητικά στοιχεία του μοντέλου	28
3.1.2	Γενίκευση του SIR με σκοπό την εφαρμογή σε αντικείμενα πάνω σε γράφους	29
3.1.3	Καταλήγοντας στην επίλυση της εξίσωσης θερμότητας	31
3.1.4	Η δυναμική εξέλιξη της εξάπλωσης - Αποτελέσματα	34
3.2	Η εξίσωση της θερμότητας πάνω σε γράφους - label propagation	37
3.2.1	Ελαχιστοποίηση με βάση τη νόρμα L_2	37
3.2.2	Ελαχιστοποίηση με βάση τη νόρμα L_1 - Αντιμετώπιση outliers	37
3.3	Αλγόριθμος Τυχαίου Περιπατητή (Random Walker)	39
3.3.1	Εισαγωγικά στοιχεία της μεθόδου	39
3.3.2	Μαθηματική διατύπωση της παραδοσιακής μεθόδου Random Walker	39
3.3.3	Σύνδεση με άλλες μεθόδους	40
3.3.4	Επεκτάσεις του Αλγορίθμου Random Walker	41
3.3.5	Περιορισμοί και πλεονεκτήματα του παραδοσιακού αλγορίθμου Random Walker	41
3.3.6	Unbiased Random Walker	43
3.3.7	Περιορισμοί και πλεονεκτήματα του Unbiased Random Walker	44
3.3.8	Οπτικοποίηση-Αποτελέσματα	45

3.3.9	Υπολογιστικό κόστος για Unbiased Random Walker	52
3.3.10	Υπολογιστικό κέρδος RAG vs Regular Grid	52
3.3.11	Συνδυάζοντας το LNP με τον αλγόριθμο Τυχαίου Περιπατητή	53
3.3.12	Regularization Framework	54
3.4	Αλγόριθμος Power Watershed	57
3.4.1	Περιγραφή του ενιαίου framework	57
3.4.2	Υλοποίηση σε μη κανονικό πλέγμα	57
3.5	Συμπεράσματα κεφαλαίου 3 - Ανακεφαλαίωση	58
4	Τεχνικές Εκμάθησης Χωρίς Επίβλεψη (Unsupervised Learning)	59
4.1	Εισαγωγή στη μέθοδο K-Means	59
4.2	Εισαγωγή στη μέθοδο NMF	60
4.2.1	Επεκτάσεις της μέθοδου NMF	61
4.2.2	Cluster NMF	61
4.3	Εφαρμογή του Cluster NMF στην κατάτμηση εικόνας	63
4.3.1	Αρχική μοντελοποίηση	63
4.3.2	Multiview Learning - συγκρίνοντας με το Cluster NMF	66
4.4	Κατασκευή - Ενσωμάτωση χωρικού όρου	68
4.4.1	Μαθηματική προσέγγιση ελαχιστοποίησης	69
4.4.2	Περιγραφή της βελτιωμένης μεθόδου	70
4.4.3	Εξαγωγή χαρακτηριστικών χρώματος	70
4.4.4	Εξαγωγή χαρακτηριστικών υφής - Gabor Filterbank	71
4.4.5	Αποτελέσματα Κατηγοριοποίησης Υφής	73
4.4.6	Αποτελέσματα Κατηγοριοποίησης Χρώματος	75
4.4.7	Συνδυάζοντας την ελαχιστοποίηση με το Regularization Framework	77
4.5	Consensus Learning	78
4.5.1	Εισαγωγή	78
4.5.2	Μαθηματική διατύπωση του Consensus Learning	78
4.5.3	Το scaling πρόβλημα και η εισαγωγή smoothness constraints - Μέρος 1	80
4.5.4	Μαθηματική επίλυση του Consensus Learning	81
4.5.5	Το scaling πρόβλημα και η εισαγωγή smoothness constraints - Μέρος 2	81
4.5.6	Color/Texture Consensus Learning	81
4.5.7	Συνολικό σύστημα	83
4.6	Συμπεράσματα κεφαλαίου 4 - Ανακεφαλαίωση	87
5	Σύγκριση μεταξύ Unsupervised και Supervised Τεχνικών	88
5.1	Τεχνικές με επίβλεψη (Supervised)	88
5.1.1	Πλεονεκτήματα Supervised τεχνικών	88
5.1.2	Μειονεκτήματα Supervised τεχνικών	88
5.2	Τεχνικές δίχως επίβλεψη (Unsupervised)	89
5.2.1	Πλεονεκτήματα Unsupervised τεχνικών	89
5.2.2	Μειονεκτήματα Unsupervised τεχνικών	89
6	Συμβολές της Διπλωματικής Εργασίας και Μελλοντική Έρευνα	91
6.1	Συμβολές της διπλωματικής εργασίας	91
6.2	Μελλοντική έρευνα	92

Κατάλογος Σχημάτων

2.1	Ο Region Adjacency γράφος (RAG).	16
2.2	Ο 8-nn γράφος.	16
2.3	Η μόνιμη κατάσταση του label-propagation για 4 διαφορετικές τιμές σ .	17
2.4	Ιδανικό clustering (αριστερά) και LNP-clustering (δεξιά), $k = 8$, 300 points.	19
2.5	Ιδανικό clustering (αριστερά) και LNP-clustering (δεξιά), $k = 5$, 300 points.	19
2.6	Σύγκριση μεταξύ LNP και GLNP, $k = 3$.	21
2.7	Ο χώρος των ισχυρών ακμών εντός του τριγώνου AOB.	22
2.8	Πιθανό \mathcal{S}_t και ο αντίστοιχος πίνακας γειτνίασης.	24
2.9	Αποτέλεσμα κατάτμησης καλής ποιότητας (house) και μέτριας (car + house + yard).	24
2.10	Ο χώρος των ισχυρών ακμών σε 3 διαστάσεις.	25
2.11	Η αρχική εικόνα και η resized εκδοχή αυτής για 5 επίπεδα νοητών κόμβων, woman.	26
2.12	Η αρχική εικόνα και η resized εκδοχή αυτής για 10 επίπεδα νοητών κόμβων, woman.	26
3.1	Κανονικό πλέγμα - regular grid.	28
3.2	Εξάπλωση ασθένειας για 100 επαναλήψεις σε μικρό/ευρύτερο κομμάτι της εικόνας, penguin.	31
3.3	Αποτελέσματα μόνιμης κατάστασης με διάφορες μεθόδους, penguin piece 1.	34
3.4	Εξάπλωση 2 ασθενειών (500 επαναλήψεις) και μόνιμη κατάσταση, penguin piece 3.	34
3.5	Εξάπλωση 2 ασθενειών και μόνιμη κατάσταση, penguin piece 3.	35
3.6	Εξάπλωση 2 ασθενειών και μόνιμη κατάσταση, woman + stairs piece.	35
3.7	Εξάπλωση 2 ασθενειών και μόνιμη κατάσταση, bird piece.	36
3.8	Αρχικές εικόνες, woman + stairs, bird.	36
3.9	Αρχικά label και τελικό L_2 και L_1 diffusion, από [9].	38
3.10	Ευαισθησία στην αρχική θέση των σπόρων, 2 classes, CT image.	42
3.11	Ευαισθησία σε outliers, 2 classes, CT image.	42
3.12	Regular 8 Grid, $\sigma_{kernel} = 90$, 2 classes, bird gray.	45
3.13	Regular 8 Grid, $\sigma_{kernel} = 90$, 3 classes, road.	45
3.14	Regular 8 Grid, $\sigma_{kernel} = 90$, 3 classes, small bird gray.	46
3.15	Regular 8 Grid, $\sigma_{kernel} = 90$, 2 classes, swan gray.	46
3.16	Ground Truth, Regular 8 Grid, $\sigma_{kernel} = 90$, 2 classes, moon.	46
3.17	Regular 8 Grid, 5 seed types, $\sigma_{kernel} \in [40, 140]$, F-score mean.	47
3.18	Regular 8 Grid, 5 seed types, $\sigma_{kernel} \in [40, 140]$, F-score standard deviation.	47
3.19	Regular 8 Grid, 10 seed types, $\sigma_{kernel} \in [1, 350]$, F-score mean.	48

3.20	Region Adjacency γράφος $\sigma_{kernel} = 90$, 2 classes, CT image.	48
3.21	Region Adjacency γράφος $\sigma_{kernel} = 90$, 2 classes, moon.	49
3.22	Region Adjacency γράφος $\sigma_{kernel} = 90$, 2 classes, women gray.	49
3.23	Region Adjacency γράφος $\sigma_{kernel} = 90$, 2 classes, boat gray.	49
3.24	Region Adjacency γράφος $\sigma_{kernel} = 90 + \text{outliers}$, 2 classes, swan gray.	50
3.25	Region Adjacency γράφος $\sigma_{kernel} = 90$, 2 classes, horses gray.	50
3.26	Region Adjacency γράφος $\sigma_{kernel} = 90$, 3 classes, castle gray.	50
3.27	Region Adjacency γράφος, Unbiased Outperforms, seeds= [1672, 1024], 2 classes, CT image.	50
3.28	Region Adjacency γράφος, Unbiased Outperforms, 2 classes, horses (using Lab).	51
3.29	Region Adjacency γράφος, Unbiased Outperforms, 2 classes, bird gray.	51
3.30	Υπολογιστικοί χρόνοι ανά χιλιάδα κόμβων.	52
3.31	Υπολογιστικοί χρόνοι ανά χιλιάδα pixel.	52
3.32	Un-normalized vs Normalized + LNP, $k = 5$, 3 classes, castle.	53
3.33	Un-normalized vs Normalized + LNP, $k = 15$, 3 classes, castle.	53
3.34	Un-normalized vs Normalized + LNP, $k = 15$, 2 classes, bird.	53
3.35	Grayscale regularization, $\sigma_{kernel} = 90$, castle gray.	55
3.36	SNR Analysis, $\sigma_{kernel} = 90$, castle gray.	55
3.37	SNR Analysis, without σ_{kernel} , castle.	56
3.38	RGB regularization, $\sigma_{kernel} = 90$, castle.	56
3.39	Watershed with Markers, Power Watershed, Random Walker, CT image	57
4.1	Ενδεικτικό αποτέλεσμα K-Means, 3 classes.	60
4.2	Εκτέλεση Cluster NMF, 3 clusters.	62
4.3	Εκτέλεση Cluster NMF, Βελτίωση κεντροειδών ανά iteration, 3 clusters.	63
4.4	Αρχικές Εικόνες, plane, swan, horses.	64
4.5	Cluster NMF vs K-Means, $k = 2$, 500 Iterations για Cluster NMF, horses.	64
4.6	Εκτέλεση Cluster NMF με τυχαία αρχικοποίηση, Iterations = [800, 1300, 1500, 1800, 3000], horses.	65
4.7	Εκτέλεση Cluster NMF $k = 3$, 500 Iterations για Cluster NMF, plane.	65
4.8	Εκτέλεση Cluster NMF $k = 2$, 500 Iterations για Cluster NMF, swan.	65
4.9	Εκτέλεση NMF και MultiView NMF με $\lambda = 0$, $k = 3$, 10 experiments, plane.	67
4.10	Εκτέλεση Cluster NMF $k = 3$, 10 experiments, plane.	67
4.11	Ισοδυναμία NMF και MultiView NMF για $\lambda = 0$, $k = 3$, plane.	68
4.12	Βήματα της βελτιωμένης μεθόδου.	70
4.13	Παράδειγμα σύγκρισης RGB με LAB, bird.	70
4.14	Παράδειγμα gabor filterbank, 4 scales and 10 orientations, με τη βοήθεια της υλοποίησης στο [32].	71
4.15	Πλάτη φιλτραρισμένων εικόνων, Brodatz 6.	72
4.16	Πείραμα με 2 classes, 30 iterations για Cluster NMF, $\lambda = [0, 5]$, Brodatz 1.	73
4.17	Πείραμα με 5 classes, 30 iterations για Cluster NMF, $\lambda = [0, 5]$, Brodatz 2.	73
4.18	Πείραμα με 2 classes, 30 iterations για Cluster NMF, $\lambda = [0, 5]$, Brodatz 3.	73
4.19	Πείραμα με 2 classes, 30 iterations για Cluster NMF, $\lambda = [0, 5]$, Brodatz 4.	74
4.20	Πείραμα με 4 classes, 30 iterations για Cluster NMF, $\lambda = [0, 5]$, Brodatz 5.	74

4.21	Πείραμα με 2 classes, 800 iterations για Cluster NMF, bad result/initialization with spatial improvement, $\lambda = [0, 5]$, Brodatz 1.	74
4.22	Πείραμα με 3 classes, 250 iterations για Cluster NMF, $\lambda = [0, 0.03]$, elephant.	75
4.23	Πείραμα με 2 classes, 250 iterations για Cluster NMF, $\lambda = [0, 0.03]$, horses.	75
4.24	Πείραμα με 3 classes, 250 iterations για Cluster NMF, $\lambda = [0, 0.03]$, castle.	75
4.25	Πείραμα με 3 classes, 250 iterations για Cluster NMF, $\lambda = [0, 0.03]$, plane.	76
4.26	Πείραμα με 2 classes, 250 iterations για Cluster NMF, $\lambda = [0, 0.03]$, eagle.	76
4.27	Πείραμα με 3 classes, 250 iterations για Cluster NMF, $\lambda = [0, 0.03]$, small bird.	76
4.28	Πείραμα με 3 classes, $NMF_{iter} = 500$, $\lambda_{spatial} = [0, 0.01]$, $\nexists \lambda_{reg}$, 2088 regions, castle.	77
4.29	Πείραμα με 3 classes, $NMF_{iter} = 500$, $\lambda_{spatial} = [0, 0.01]$, $REG_{iter} = 100$, $\sigma_{kernel} = 90$, $\lambda_{reg} = 0.01$, 2041 regions, castle.	77
4.30	Πείραμα με 3 classes, $NMF_{iter} = 500$, $\lambda_{spatial} = [0, 0.01]$, $REG_{iter} = 100$, $\nexists \sigma_{kernel}$, $\lambda_{reg} = 15$, 2008 regions, castle.	77
4.31	Color Segmentation (pixel + node level), αρχικά rank: 3, 4, 5, 6 τελικό rank: 3, $\lambda_{base} = 0.02$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, bird.	82
4.32	Texture Segmentation (pixel + node level), αρχικά rank: 2, 3, 4, 5, τελικό rank: 3, $\lambda_{base} = 0.02$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, Brodatz 4.	83
4.33	Συνολικό σύστημα (pixel level, αρχικά rank: 3(Y), 4(Y), 3(X), 4(X), τελικό rank: 2, $\lambda_{con} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = 0$, horses, zebra, grass.	84
4.34	Συνολικό σύστημα και μεταβλητότητα αποτελέσματος (pixel level, αρχικά rank: 3(Y), 4(Y), 3(X), 4(X), τελικό rank: 2, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = 0$, eagle.	84
4.35	Συνολικό σύστημα και μεταβλητότητα αποτελέσματος (pixel level, αρχικά rank: 3(Y), 4(Y), 3(X), 4(X), τελικό rank: [2,3,4], $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = 1$, kangaroo.	84
4.36	Δύσκολες περιπτώσεις (pixel level, αρχικά rank: 3(Y), 4(Y), 3(X), 4(X), τελικό rank: 2, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = 1$, kangaroo, tiger, small bear.	85
4.37	Βελτίωση λόγω χωρικού όρου λ_{con} σε χρώμα μόνο (pixel level, αρχικά rank: 3(X), 4(X), 3(X), 4(X), τελικό rank: 2, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, women.	85
4.38	Βελτίωση λόγω χωρικού όρου λ_{con} (pixel level, αρχικά rank: 3(Y), 4(Y), 3(X), 4(X), τελικό rank: 3, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, small bird.	85
4.39	Πρόσθετο αποτέλεσμα (pixel level, αρχικά rank: 3(Y), 4(Y), 3(X), 4(X), τελικό rank: 5, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, woman + hair.	85
4.40	Σχεδιάγραμμα συνολικού συστήματος.	86
5.1	Σύγκριση (pixel level), $\sigma_{RW} = 90$, αρχικά rank: 3(X), 4(X), 5(X), 6(X), τελικό rank: 3, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, bird.	90

5.2	Σύγκριση (pixel level), $\sigma_{RW} = 100$, αρχικά rank: 3(X), 4(X), 5(X), 6(X), τελικό rank: 3, $\lambda_{base} = 0.02$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, bird.	90
5.3	Σύγκριση (pixel level), $\sigma_{RW} = 90$, αρχικά rank: 5(X), 6(X), 5(X), 6(X), τελικό rank: 6, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, boatman.	90

Κεφάλαιο 1

Εισαγωγή

1.1 Η πληροφορία στη σύγχρονη εποχή

Η σύγχρονη εποχή χαρακτηρίζεται κυρίως από την έννοια της πληροφορίας καθώς και το μέγεθος αυτής. Χαρακτηριστικό παράδειγμα αποτελεί το Διαδίκτυο και ιδιαιτέρως τα κοινωνικά δίκτυα, όπου κάθε μέρα εκατομμύρια χρηστών μοιράζονται πληροφορίες σε μορφή εικόνας/video, αλλά και πολυμεσικές εφαρμογές όπως το Youtube ή το Flickr. Για να κατανοήσουμε και να εκμεταλλευτούμε αυτήν την τεράστια ποσότητα δεδομένων (η οποία πλέον έχει καθιερωθεί και με τον όρο Big Data), απαιτείται η αποθήκευση αλλά και η καθιέρωση εξελιγμένων μεθόδων ανάκτησης και επεξεργασίας αυτής. Παράλληλα, η αποτελεσματική συλλογή πληροφορίας καθίσταται απαραίτητη ώστε τα σύγχρονα υπολογιστικά μηχανήματα να μπορέσουν να εξάγουν χρήσιμα συμπεράσματα για τους χρήστες τους.

Πέραν της εγγενούς δυσκολία διαχείρισης του όγκου της πληροφορίας, τίθεται συχνά και το ζήτημα της δομής των δεδομένων. Για παράδειγμα, σε ένα κοινωνικό δίκτυο, κάθε χρήστης μπορεί να έχει εκατοντάδες έως και λίγες χιλιάδες επαφών. Με τη βοήθεια του παραδείγματος αυτού, καταλαβαίνουμε την πολυπλοκότητα την οποία μπορούν να αποκτήσουν τα δεδομένα, η οποία μας αναγκάζει να εισάγουμε σύνθετα μαθηματικά μοντέλα για την ανάλυση τους. Ένα πολύ χρήσιμο μαθηματικό εργαλείο με πληθώρα εφαρμογών και αντίστοιχης επιστημονικής έρευνας είναι και οι γράφοι.

Ως γράφος θεωρείται ένα σύνολο κόμβων οι οποίοι συνδέονται μεταξύ τους με ακμές. Με τον τρόπο αυτό μοντελοποιούνται σχέσεις μεταξύ των κόμβων, τους οποίους μπορούμε αφαιρετικά να θεωρήσουμε σα γενικευμένες οντότητες. Για παράδειγμα, το κοινωνικό δίκτυο ενός χρήστη στο διαδίκτυο μπορεί να αναπαρασταθεί με κόμβους για τους χρήστες και ακμές για τον χαρακτηρισμό αυτών (για παράδειγμα πόσο συχνά μοιράζονται πληροφορίες ή τον τύπο των δεδομένων τα οποία ανταλλάσσουν).

1.2 Όραση Υπολογιστών και γραφοθεωρητικές μέθοδοι

Ένα από τα πιο παραδοσιακά προβλήματα στην Όραση Υπολογιστών είναι η ανίχνευση αντικειμένων πάνω σε εικόνες/video. Στα πλαίσια αυτής της κατεύθυνσης, υπάρχει εκτενής έρευνα πάνω σε εφαρμογές όπως η κατάτμηση εικόνας, η ανίχνευση των ακμών, η κατηγοριοποίηση υφής και πολλές άλλες. Κοινός παρονομαστής αυτών των εφαρμογών είναι η θεμελίωση αλγοριθμικών μεθόδων οι οποίες χρησιμοποιούν τις τιμές φωτεινότητας των pixels μιας εικόνας και τις θέσεις αυτών, ώστε με κατάλληλες τεχνικές να εξαχθούν τα αντικείμενα τα οποία υπάρχουν σε μια εικόνα.

Τα τελευταία χρόνια, ο όγκος των δεδομένων και η ανάγκη χειρισμού αυτών γρήγορα (όπως σε βιοϊατρικές εφαρμογές), οδήγησε στην καθιέρωση γραφοθεωρητικών μεθόδων στην Όραση Υπολογιστών. Αυτές οι μέθοδοι επεκτείνουν την παραδοσιακή έννοια του κανονικού πλέγματος σε μια εικόνα σε μη κανονικούς γράφους οι οποίοι αναπαριστούν πιο πολύπλοκες δομές. Με τον τρόπο αυτό, η Όραση Υπολογιστών εμπλουτίστηκε με ιδέες από άλλα επιστημονικά πεδία και ταυτόχρονα άνοιξε το δρόμο για περισσότερες δυνατότητες, μέσα από αυτήν την εξέλιξη.

1.3 Κεντρική ερευνητική κατεύθυνση και συμβολές της διπλωματικής εργασίας

Βασική επιδίωξη της εργασίας είναι η χρήση γράφων για την αναπαράσταση και ανάλυση των εικόνων με τελικό σκοπό την κατάτμηση σε περιοχές ενδιαφέροντος. Συγκεκριμένα, στο κεφάλαιο 2, αφού γίνει μια παρουσίαση των μαθηματικών εργαλείων που απαιτούνται, επιχειρείται η ανάδειξη σημαντικών δομών σε ένα γράφο δανειζόμενοι ιδέες από την επιστήμη των Δικτύων Υπολογιστών (Network Science). Στη συνέχεια, στο κεφάλαιο 3, παρουσιάζεται το επιδημιολογικό μοντέλο SIR και αποδεικνύεται η σύνδεση του με τεχνικές διάχυσης θερμότητας πάνω σε γράφους. Με τη βοήθεια αυτής της σύνδεσης, εισάγεται μια νέα αλγοριθμική μέθοδος η οποία και εφαρμόζεται σε πραγματικά δεδομένα εικόνων.

Στο κεφάλαιο 4, εισάγεται ένα νέο σύστημα κατάτμησης εικόνων δίχως επίβλεψη, το οποίο βασίζεται σε ιδέες από Εκμάθηση Μηχανών (Machine Learning). Στο κεφάλαιο 5, γίνεται μια σύγκριση μεταξύ των τεχνικών δίχως επίβλεψη σε σχέση με τις τεχνικές με επίβλεψη, χρησιμοποιώντας τη γνώση η οποία αντλήθηκε από τα προηγούμενα κεφάλαια. Τέλος, στο σύντομο κεφάλαιο 6, παρουσιάζονται τα βασικά σημεία της εργασίας με συμπυκνωμένο τρόπο και καταγράφονται μελλοντικές ερευνητικές κατευθύνσεις.

Κεφάλαιο 2

Κατασκευή Γράφων και Ρύθμιση Παραμέτρων

2.1 Εισαγωγή στους γράφους

Θεωρούμε ένα γράφο $\mathcal{G}(\mathcal{V}, \mathcal{E})$ όπου \mathcal{V} είναι το σύνολο των κόμβων του \mathcal{G} και \mathcal{E} το σύνολο των ακμών του. Οι κόμβοι του συνδέονται μεταξύ τους με ακμές e_{ij} όπου i και j οι προσκείμενοι στην ακμή κόμβοι κάθε φορά. Για την περίπτωση ενός γράφου με βάρη ορίζουμε τον πίνακα $W = [w_{ij}]$ όπου w_{ij} επιλέγονται μη αρνητικά ώστε να συμβολίζουν την ομοιότητα μεταξύ των κόμβων i και j . Για τους σκοπούς των εφαρμογών που μας απασχολούν, θεωρούμε ότι ο \mathcal{G} είναι ένας συνεκτικός γράφος με βάρη. Επίσης, υποθέτουμε ότι η σχέση ομοιότητας μεταξύ 2 κόμβων είναι σχέση συμμετρική, δηλαδή ο πίνακας W είναι συμμετρικός ($W = W^T$). Με άλλα λόγια θεωρούμε ότι πέραν της συνεκτικότητας του γράφου και της ύπαρξης ακμών με βάρη, ο γράφος είναι μη κατευθυνόμενος.

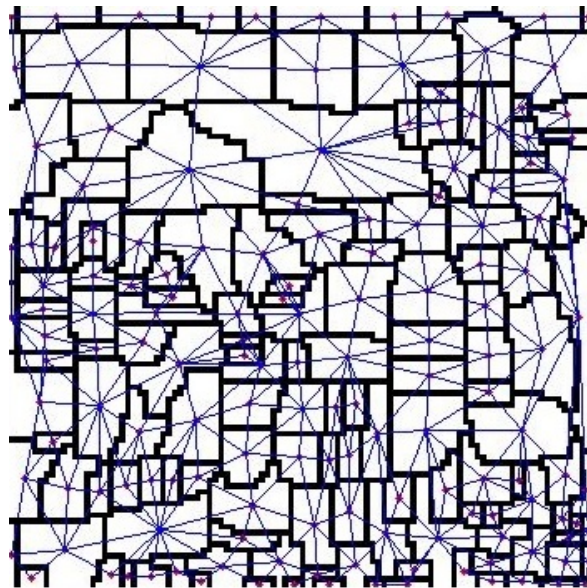
Για εφαρμογές που αντιστοιχούν σε εικόνες/video συνηθίζεται να αναπαρίσταται μια εικόνα g ως ένας γράφος \mathcal{G} με κόμβους οι οποίοι τοποθετούνται χωρικά στις θέσεις των pixel της εικόνας. Κατ' αναλογία, ορίζεται το σύνολο των ακμών ώστε να συνδέεται κάθε pixel με τα αντίστοιχα 4 ή 8 γειτονικά του. Στην περίπτωση αυτή θα αναφερόμαστε με τον όρο 4/8 regular grid. Προφανώς, η πληθικότητα (cardinality) του συνόλου των κόμβων $|\mathcal{V}|$ ισούται με $n \times m$ όπου n και m είναι η διάσταση της εικόνας στις διευθύνσεις x και y αντίστοιχα.

Εναλλακτικά, μπορούμε να εφαρμόσουμε το μετασχηματισμό watershed στη grayscale εκδοχής της εικόνας g και να λάβουμε μια υπερκατατημημένη εκδοχή αυτής ([17]) υπό μορφή περιοχών στις οποίες ανήκουν συνήθως περισσότερα του ενός pixel. Οι περιοχές αυτές διαχωρίζονται μεταξύ τους μέσω των boundaries που έχουν προκύψει από το μετασχηματισμό. Στην περίπτωση αυτή, η πληθικότητα (cardinality) του συνόλου των κόμβων $|\mathcal{V}|$ ισούται με n , όπου n είναι ο αριθμός των περιοχών που προέκυψαν.

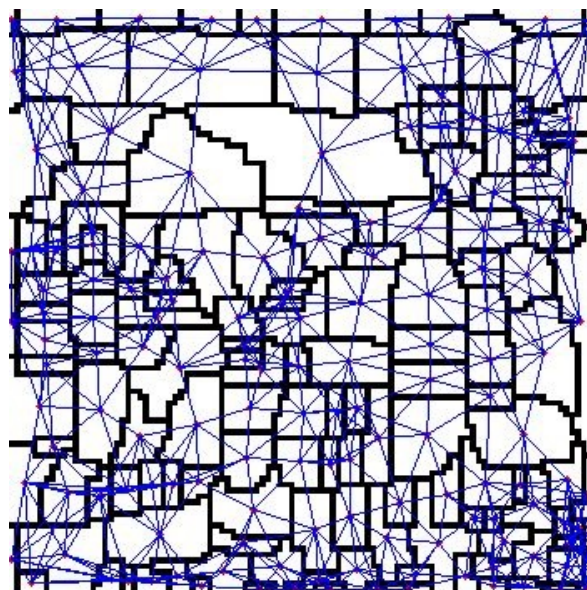
Έπειτα, ορίζουμε τα γεωμετρικά κεντροειδή $x_i = (x_{i,1} \ x_{i,2})^T$ των περιοχών αυτών σαν τις χωρικές συντεταγμένες των κόμβων, όπου i είναι ο δείκτης του κόμβου ή της περιοχής που εξετάζουμε. Επιπλέον, υπολογίζουμε τη μέση τιμή της φωτεινότητας των pixel εντός της περιοχής την οποία συμβολίζουμε με g_i . Σημειώνεται ότι για grayscale εικόνες το g_i είναι βαθμωτή ποσότητα, ενώ για έγχρωμες εικόνες είναι ένα διάνυσμα με περισσότερες

συνιστώσες. Για παράδειγμα, στο χρωματικό χώρο RGB θα έχουμε $g_i = (g_{i,1} \ g_{i,2} \ g_{i,3})^\top$.

Αφότου καθοριστούν οι κόμβοι του γράφου, απαιτείται ο προσδιορισμός των ακμών. Δύο συνηθισμένες επιλογές είναι ο Region Adjacency γράφος (RAG) και ο k -nearest neighborhood (k -nn) γράφος. Στην πρώτη περίπτωση, περιοχές οι οποίες μοιράζονται κοινά pixel boundaries, συνδέονται με ακμή οι οποίοι συνδέει τους κόμβους των αντίστοιχων περιοχών. Με τον τρόπο αυτό, ο αριθμός των εξερχόμενων ακμών από μια περιοχή δεν είναι σταθερός. Στη δεύτερη περίπτωση, κάθε περιοχή συνδέεται με τους k κοντινότερους γείτονες, δηλαδή τους k κόμβους οι οποίοι βρίσκονται πιο κοντά στον κόμβο της εκάστοτε περιοχής. Προφανώς, στην περίπτωση αυτή, μπορεί γειτονικές περιοχές να μη συνδέονται με ακμή.



Σχήμα 2.1: Ο Region Adjacency γράφος (RAG).



Σχήμα 2.2: Ο 8-nn γράφος.

2.2 Κατασκευή γράφου για εφαρμογές ανίχνευσης αντικειμένων

2.2.1 Τυπικές συναρτήσεις ομοιότητας

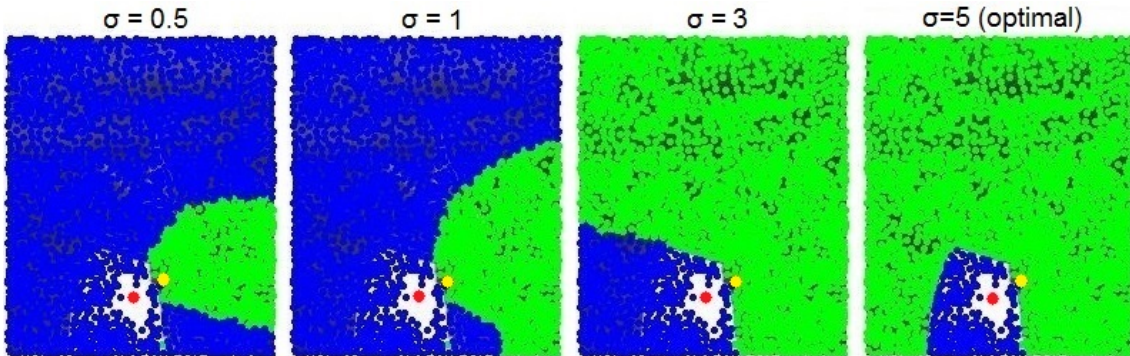
Εστιάζουμε τώρα σε εφαρμογές πάνω σε εικόνες. Η κατασκευή του πίνακα $W = [w_{i,j}]$ είναι ένα από τα πιο σημαντικά στάδια καθώς μοντελοποιεί τις σχέσεις ομοιότητας μεταξύ των κόμβων του γράφου. Στη γενική περίπτωση όπου δεν επιλέγουμε 4/8 regular grid, μπορεί να θεωρηθεί ότι κάθε κόμβος του γράφου συνδέεται με τους k κοντινότερους γείτονες του. Εναλλακτικά, μπορεί να χρησιμοποιηθεί ένας region adjacency γράφος, όπου μόνο περιοχές προσκείμενες η μία στην άλλη συνδέονται μεταξύ τους με ακμή μη μηδενικού βάρους. Εν γένει, τα βάρη αυτά ενσωματώνουν την ομοιότητα μεταξύ δύο κόμβων i και j με ένα RBF (Radial Basis Function) kernel της μορφής:

$$w_{i,j} = \begin{cases} \exp \left[- \left(\frac{\|g_i - g_j\|_2}{\sigma_g} \right)^2 \right] & \text{αν } j \sim i \\ 0 & \text{αλλιώς} \end{cases} \quad (2.1)$$

Ο όρος σ_g καθορίζει το εύρος της συνάρτησης του $w_{i,j}$, ενώ ο συμβολισμός $j \sim i$ δηλώνει ότι ο κόμβος j είναι γείτονας του κόμβου i . Για τους σκοπούς της εκάστοτε εφαρμογής, ο προσδιορισμός του γίνεται manually είτε κάποιες φορές με τεχνικές cross-validation. Εναλλακτικά, έχει προταθεί και ο τοπικός υπολογισμός της παραμέτρου ([11]). Στη βιβλιογραφία ([17],[16]) έχουν προταθεί παρόμοιες συναρτήσεις, οι οποίες μπορεί να περιλαμβάνουν χωρικούς όρους στο kernel της μορφής

$$w_{i,j} = \begin{cases} \exp \left[- \left(\left(\frac{\|g_i - g_j\|_2}{\sigma_g} \right)^2 + \left(\frac{\|x_i - x_j\|_2}{\sigma_x} \right)^2 \right) \right] & \text{αν } j \sim i \\ 0 & \text{αλλιώς} \end{cases} \quad (2.2)$$

Με το ακόλουθο πείραμα, βλέπουμε ότι μικρές αλλαγές στην τιμή του σ (θα αναφερόμαστε στο σ_g με τον όρο σ ή σ_{kernel} όταν δε χρησιμοποιείται ο όρος σ_x) επιφέρουν δραστικές αλλαγές στο αποτέλεσμα της κατάτμησης. Στο παράδειγμα αυτό, χρησιμοποιούμε έναν region adjacency γράφο και την (2.1) για το σχηματισμό του πίνακα W . Έπειτα, εισάγουμε δύο σημεία - σπόρους και υλοποιούμε τη μέθοδο του label propagation (βλέπε κεφάλαιο 3). Το βέλτιστο αποτέλεσμα αντιστοιχεί στην τιμή $\sigma = 5$.



Σχήμα 2.3: Η μόνη κατάσταση του label-propagation για 4 διαφορετικές τιμές σ .

2.2.2 Μέθοδος Linear Neighborhood Propagation (LNP)

Σε πολλές εφαρμογές, η σωστή επιλογή της παραμέτρου σ μπορεί να είναι υπολογιστικά δύσκολη και χρονοβόρα, διότι μπορεί να απαιτεί πρόσθετους υπολογισμούς ή επανεκτελέσεις της μεθόδου. Για το λόγο αυτό, προτάθηκε στο [1] η μέθοδος του Linear Neighborhood Propagation (LNP) ([1], [2]). Η πρωταρχική ιδέα της μεθόδου είναι ότι η τιμή x_i σε έναν κόμβο i ενός feature vector x μπορεί να προσεγγιστεί από ένα σταθμισμένο γραμμικό συνδυασμό των τιμών που λαμβάνουν οι άμεσοι γείτονες του κόμβου (έστω $N(x_i)$ η γειτονιά του κόμβου i), δηλαδή:

$$\hat{x}_i = \sum_{j: x_j \in N(x_i)} w_{i,j} x_j \quad (2.3)$$

Στη συνέχεια, μπορεί κανείς να ελαχιστοποιήσει την απόσταση μεταξύ της προσέγγισης \hat{x}_i και της πραγματικής τιμής x_i ώστε να πάρει το βέλτιστο δυνατό διάνυσμα w_i το οποίο περιέχει τις τιμές βάρους του κόμβου i με τους γείτονες του. Έπειτα, αθροίζοντας αυτές τις ευκλείδειες αποστάσεις λάθους παίρνει κανείς το $\epsilon(W)$ το οποίο αποτελεί το συνολικό reconstruction error του \hat{X} με τον X οι οποίοι περιέχουν όλα τα feature vector των κόμβων.

$$\epsilon(W) = \sum_i \epsilon_i = \sum_i \|x_i - \hat{x}_i\|^2 \quad (2.4)$$

Η ελαχιστοποίηση της ποσότητας ϵ_i συνοδεύεται από τους περιορισμούς

$$\sum_{j: x_j \in N(x_i)} w_{i,j} = 1 \quad (2.5)$$

$$w_{i,j} = 0 \quad \forall i \neq j \quad (2.6)$$

Η (2.5) εξασφαλίζει ότι οι γραμμές του W αθροίζουν στη μονάδα, ενώ η (2.6) το sparsity του πίνακα. Η ελαχιστοποίηση της (2.4) προκύπτει με βάση την παρατήρηση ότι:

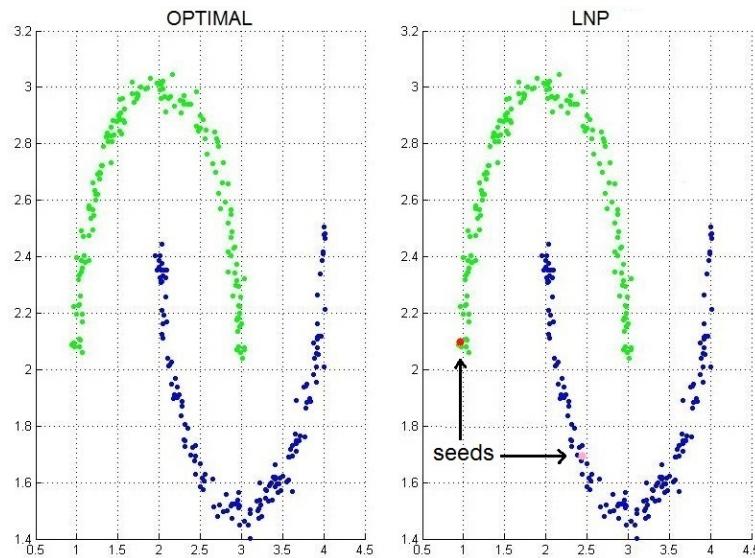
$$\epsilon_i = \left\| x_i - \sum_{j: x_j \in N(x_i)} w_{i,j} x_j \right\|^2 = \left\| \sum_{j: x_j \in N(x_i)} w_{i,j} (x_i - x_j) \right\|^2$$

$$\epsilon_i = \sum_{j, k: x_j, x_k \in N(x_i)} w_{i,j} w_{i,k} (x_i - x_j)^\top (x_i - x_k) = \sum_{j, k: x_j, x_k \in N(x_i)} w_{i,j} G_{j,k}^i w_{i,k}$$

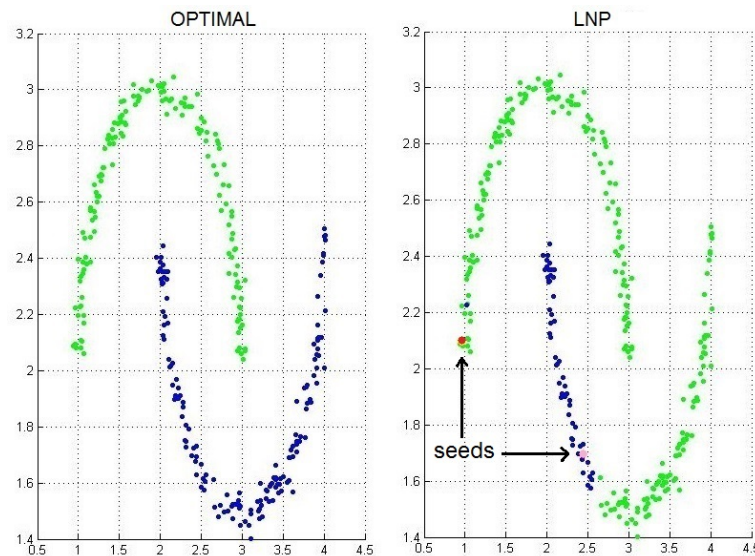
όπου G^i δηλώνει τον κόμβο i γύρω από τον οποίο κεντράρεται ο υπολογισμός του $(x_i - x_j)^\top (x_i - x_k)$. Δεδομένου ότι έχουμε $\epsilon_i \quad \forall i = 1, \dots, n$ απαιτείται να επιλυθούν n τέτοιες ελαχιστοποιήσεις, δηλαδή n τετραγωνικά προβλήματα:

$$\begin{aligned} \min_{w_{i,j}} \quad & \sum_{j, k: x_j, x_k \in N(x_i)} w_{i,j} G_{j,k}^i w_{i,k} \\ \text{s.t.} \quad & \sum_j w_{ij} = 1, w_{i,j} \geq 0 \end{aligned}$$

Η ποσότητα $N(x_i)$ τίθεται ίση με τους k κοντινότερους γείτονες για τον x_i , δηλαδή χρησιμοποιείται $k - nn$ γράφος. Από τη στιγμή που έχει δημιουργηθεί ο sparse πίνακας W είναι δυνατό το label propagation επί του γράφου \mathcal{G} ([1], [5]). Ένα από τα κυριότερα προβλήματα της μεθόδου είναι ότι πρέπει να χρησιμοποιηθεί συγκεκριμένο k γεγονός το οποίο οδηγεί σε προβλήματα γενίκευσης. Όταν το k είναι σχετικά μικρό, εμφανίζονται μη συνεκτικά κομμάτια στο γράφο \mathcal{G} , ενώ για πολύ μεγαλύτερες τιμές συνδέονται σημεία τα οποία ανήκουν σε διαφορετικές κλάσεις ([3]). Ένα ακόμη μειονέκτημα του LNP είναι ότι ο πίνακας W ο οποίος προκύπτει δεν είναι συμμετρικός εκ κατασκευής (αφού τα n προβλήματα είναι ανεξάρτητα μεταξύ τους), οπότε χρειάζεται να τον τροποποιήσουμε ορίζοντας μια συμμετρική εκδοχή $\tilde{W} = [\tilde{w}_{i,j}]$ του W ώστε $\tilde{w}_{i,j} = \frac{1}{2}(w_{i,j} + w_{j,i})$, τροποποίηση η οποία όταν η ποσότητα $|w_{i,j} - w_{j,i}|$ είναι μεγάλη δεν είναι ακριβής.



Σχήμα 2.4: Ιδανικό clustering (αριστερά) και LNP-clustering (δεξιά), $k = 8$, 300 points.



Σχήμα 2.5: Ιδανικό clustering (αριστερά) και LNP-clustering (δεξιά), $k = 5$, 300 points.

2.2.3 Μέθοδος Global Linear Neighborhood Propagation (GLNP)

Επειδή το classification performance της προηγούμενη μεθόδου εξαρτάται κυρίως από την επιλογή της παραμέτρου k , έχει προταθεί η μέθοδος Global Linear Neighborhood Propagation (GLNP) η οποία στηρίζεται όχι μόνο στους άμεσα γείτονες ενός κόμβου αλλά και στους γείτονες αυτών κοκ ([3]). Στην περίπτωση του GLNP, θεωρούμε ότι το x_i ανακατασκευάζεται προσεγγιστικά από τον εαυτό του, τους άμεσα γείτονες του, τους άμεσα γείτονες αυτών κλπ, με την επίδραση τους κάθε φορά να ελαττώνεται κατά έναν παράγοντα $a \in [0, 1]$. Πιο συγκεκριμένα:

$$\hat{X} = (1 - a)(X + aSX + a^2S^2X + \dots) = (1 - a) \sum_{i=0}^{\infty} (aS)^i X, S = D^{-\frac{1}{2}}WD^{-\frac{1}{2}} \quad (2.7)$$

με $D = [d_i] = \left[\sum_{j \sim i} w_{ij} \right]$ διαγώνιο $n \times n$ πίνακα. Για $i = 1$ προκύπτει ο όρος του LNP, δηλαδή το GLNP αποτελεί γενίκευση της τοπικής προσέγγισης του LNP. Παρατηρούμε ακόμα ότι όταν οι ιδιοτιμές του aS είναι κατ' απόλυτο τιμή μικρότερες της μονάδας, τότε:

$$\sum_{i=0}^{\infty} (aS)^i X = (I - aS)^{-1} X$$

μπορούμε λόγω της (2.7) να ελαχιστοποιούμε την ποσότητα

$$E(S) = \|X - (1 - a)(I - aS)^{-1}X\|^2 \quad (2.8)$$

όπου

$$\|A - B\|^2 = \sum_{ij} (A_i - B_j)^2$$

Σημειώνεται ότι ο X είναι πίνακας $n \times m$ όπου n ο αριθμός των κόμβων και m η διάσταση του feature vector που έχουμε επιλέξει. Η βασική ιδέα για τη λύση της ελαχιστοποίησης της (2.8) είναι η εξής ([3]): ο πίνακας S είναι συμμετρικός με πραγματικές τιμές στοιχείων. Επομένως, μπορούμε να τον γράψουμε ως

$$S = Q\Lambda Q^T, Q \text{ ορθοκανονικός και } \Lambda \text{ διαγώνιος με πραγματικές τιμές}$$

$$E(S) = \|X - (1 - a)(I - aQ\Lambda Q^T)^{-1}X\|^2 = \|X - (1 - a)Q(I - a\Lambda)^{-1}Q^T X\|^2$$

Στη συνέχεια, μπορούμε να κρατήσουμε τις k κύριες (μεγαλύτερες) ιδιοτιμές του S το οποίο ισοδυναμεί τελικά με το να κρατήσουμε από τον $(I - a\Lambda)^{-1}$ τις k μεγαλύτερες τιμές στη διαγώνιο του (πίνακας $(I - a\Lambda)_{(k)}^{-1}$ διάστασης $k \times k$) και τις αντίστοιχες στήλες από τον Q (πίνακας $Q_{(k)}$ διάστασης $n \times k$). Τελικά, ορίζουμε

$$F = \sqrt{1 - a}Q_{(k)}(I - a\Lambda)_{(k)}^{-1}$$

και προσεγγίζουμε το $E(S) \approx \|X - FF^T X\|^2$, έχοντας προσεγγίσει τον S ως εξής:

$$S \approx Q_{(k)}\Lambda_{(k)}Q_{(k)}^\top = F \frac{(I - a\Lambda)^{\frac{1}{2}}_{(k)}}{\sqrt{1-a}} \Lambda_{(k)} \frac{(I - a\Lambda)^{\frac{1}{2}}_{(k)}}{\sqrt{1-a}} F^\top = F \frac{(I - a\Lambda)_{(k)}}{1-a} \Lambda_{(k)} F^\top$$

Ο πίνακας F είναι $n \times k$ όπου k είναι ο αριθμός των (μεγαλύτερων) συνιστωσών που χρησιμοποιήθηκαν. Αν υποθεθεί ότι ο πίνακας X έχει μη αρνητικές τιμές, τότε μπορούμε να επικαλεστούμε μια παρόμοια ανάλυση με εκείνη του Non-Negative Matrix Factorization (NMF) - [6] και προκύπτει:

$$\min_{F \geq 0} Q(F) = \|X - FF^\top X\|^2 \quad (2.9)$$

Μπορούμε να επιλύσουμε τη (2.9) με gradient descent:

$$\frac{\partial Q(F)}{\partial F} = -2(X - FF^\top X)X^\top F - 2X(X^\top - X^\top FF^\top)F$$

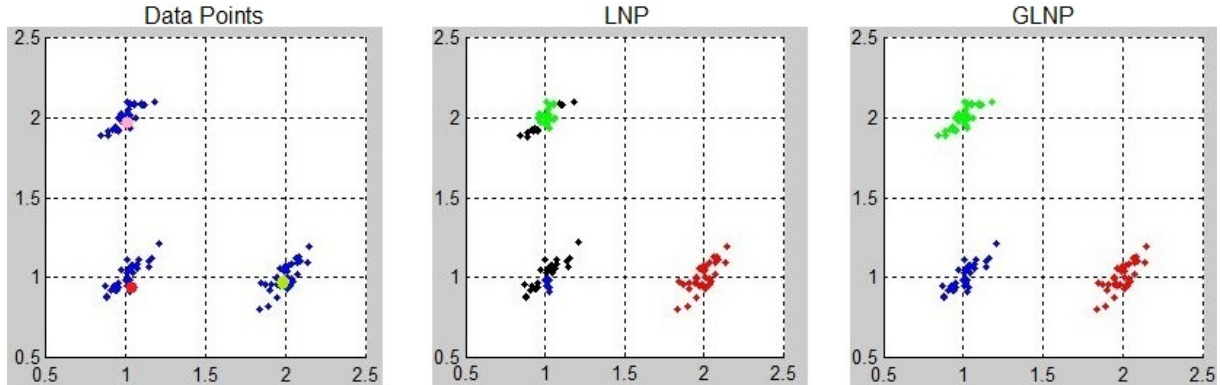
οπότε ορίζουμε αντίστοιχα μια επαναληπτική λύση για τον F :

$$F_{i,j}^{t+1} = F_{i,j}^t - \eta_{i,j} \frac{\partial Q(F_{i,j}^t)}{\partial F_{i,j}^t} = F_{i,j}^t - \eta_{i,j} [(X - FF^\top X)X^\top F + X(X^\top - X^\top FF^\top)F]_{i,j}$$

με $\eta_{i,j} > 0$ να είναι το learning rate. Αν επιπλέον υποθέσουμε ότι τα data στον πίνακα Q είναι μη αρνητικά μπορούμε ([3]) να πάρουμε μια επαναληπτική λύση της μορφής:

$$F_{i,j}^{t+1} = F_{i,j}^{t+1} \sqrt{\frac{[2XX^\top F]_{i,j}^t}{[FF^\top XX^\top F + XX^\top FF^\top F]_{i,j}^t}}$$

Στη συνέχεια, ο πίνακας $\mathcal{W} = FF^\top$ είναι συμμετρικός και κάθε στοιχείο του εκφράζει την ομοιότητα του i κόμβου με τον j . Επομένως, μπορεί κανείς να χρησιμοποιήσει τον πίνακα \mathcal{W} στο label propagation ώστε να οδηγηθεί στο τελικό data segmentation. Με βάση τα παραπάνω, μπορούμε να συγκρίνουμε μεταξύ τους τις μεθόδους LNP και GLNP. Για ισονομία μεταξύ των 2 μεθόδων θέτουμε $k = 3$ ώστε να αντιστοιχίσουμε τον αριθμό των clusters με το βαθμό της λύσης για το GLNP. Προφανώς, λόγω μικρού k πολλά σημεία στο LNP δε μπορούν να ταξινομηθούν, διότι σχηματίζονται μικρότερα υπο-clusters τα οποία δε συνδέονται μεταξύ τους.



Σχήμα 2.6: Σύγκριση μεταξύ LNP και GLNP, $k = 3$.

2.3 Αυτοματοποιημένη ρύθμιση παραμέτρων

2.3.1 Περιγραφή της βασικής ιδέας

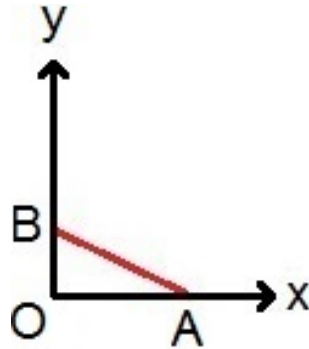
Στις προηγούμενες ενότητες, παρουσιάστηκαν 2 μέθοδοι οι οποίες κατασκευάζουν έναν πίνακα ομοιότητας μεταξύ των κόμβων του γράφου, με σκοπό την περαιτέρω χρήση του σε τεχνικές οι οποίες ανιχνεύουν clusters εντός του γράφου. Επομένως, η αποτελεσματική κατάτμηση ενός γράφου (και της αντίστοιχης εικόνας) προαπαιτεί την ποσοτικοποίηση της ομοιότητας μεταξύ των κόμβων και την απόδοση αντίστοιχων βαρών στις αντίστοιχες ακμές. Για το λόγο αυτό, μπορεί κανείς να προσπαθήσει να ανακαλύψει αυτόν το συσχετισμό μέσα από τη δομή της βασικής συνάρτησης ομοιότητας. Πιο συγκεκριμένα, έχουμε ισχυρή ακμή όταν η συνάρτηση $f(u, v)$ για 2 κόμβους u, v υπερβαίνει ένα μη αρνητικό κατώφλι ομοιότητας t , δηλαδή:

$$f(u, v) > t \Leftrightarrow \exp\left(-\left(\|d(u, v)\|^2 + \|i(u, v)\|^2\right)\right) > t \quad (2.10)$$

$$\stackrel{\ln \uparrow}{\Leftrightarrow} -(\|d(u, v)\|^2 + \|i(u, v)\|^2) > \ln(t) \quad (2.11)$$

$$\Leftrightarrow \|d(u, v)\|^2 + \|i(u, v)\|^2 < \ln\left(\frac{1}{t}\right) \quad (2.12)$$

οπότε, αν $A = \|d(u, v)\|^2$ (όρος αποστάσεων μεταξύ κόμβων), $B = \|i(u, v)\|^2$ (όρος φωτεινότητας) και $t_{new} = \ln\left(\frac{1}{t}\right)$, η (2.12) γίνεται $A + B < t_{new}$. Απαιτείται, δηλαδή, η εύρεση των συντεταγμένων ώστε να διαπιστώσουμε αν βρισκόμαστε εντός της περιοχής που ορίζει η ευθεία $A + B = t_{new}$. Για όλα αυτά τα σημεία εντός θεωρούμε ότι η ακμή είναι ισχυρή και οι προσκείμενοι σε αυτήν κόμβοι είναι ισχυρά συνδεδεμένοι.



Σχήμα 2.7: Ο χώρος των ισχυρών ακμών εντός του τριγώνου AOB.

Θεωρούμε την 4άδα των συντεταγμένων x, y, x_s, y_s . Τα x, y αντιστοιχούν στις τιμές που αντιστοιχούν στην εκάστοτε ακμή, και έστω ότι το x αφορά την απόσταση (υψώμενη στη δεύτερη δύναμη) και το y αφορά τη διαφορά φωτεινότητας (υψώμενη στη δεύτερη δύναμη πάλι) που υπολογίζονται κάθε φορά. Για το σημείο $(x, 0)$ αν θεωρηθεί $x = x_s$ υπάρχει ένα y_s το οποίο δίνεται από τη γραμμική σχέση $y_s = -\frac{B_s}{A_s}x_s + B_s$ και προκύπτει το σημείο (x_s, y_s) το οποίο ανήκει στην ευθεία (ε) που ορίζεται από την $A + B = t_{new}$. Έπειτα, συγκρίνουμε τα y, y_s μεταξύ τους. Προφανώς, προκύπτει η συνθήκη:

$$edge = \begin{cases} strong & \text{αν } y < y_s \\ weak & \text{αλλιώς} \end{cases}$$

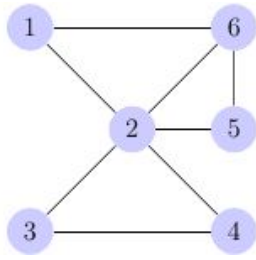
Ο όρος αυτοματοποιημένο κατώφλι έγκειται στο ότι τα A, B δεν τίθενται με κάποια τιμή γενικευμένη για όλες τις ακμές σε κάθε εικόνα αλλά ορίζονται σε κάθε ακμή ξεχωριστά. Η πιο διαισθητική επιλογή είναι να διαλέξει κανείς σαν A το μέσο όρο του μέσου όρου των αποστάσεων του κάθε κόμβου (από όλους του τους γείτονες) που συμμετέχει στην ακμή, από τη στιγμή που η σχέση ακμής είναι συμμετρική (και αντίστοιχα για το B). Η επιλογή του να παίρνει κανείς το μέσο όρο και για τις αποστάσεις του κάθε κόμβου από τους γείτονες γίνεται ώστε να περιορίζονται συνδέσεις με ακμές πολύ μακρινές και να προτιμώνται οι κοντινότερες. Αντίστοιχη επιλογή γίνεται για το B, αν αντί για απόσταση πάρουμε τις διαφορές των φωτεινότητων, όπως έχουν οριστεί εξαρχής. Για τις αρχικές ακμές του γράφου μπορούμε να επιλέξουμε $k - nn$ με $k = 10$.

Με τον τρόπο αυτό, έχουμε αποφανθεί για τις ισχυρές ακμές οι οποίες συνδέουν τους κόμβους του γράφου. Για να κάνουμε την τελική κατάτμηση της εικόνας, βρίσκουμε τα p συνεκτικά υποσύνολα με πληθικό αριθμό c_t , $t = 1..p$ αυτού του συνόλου ακμών, δηλαδή τα σύνολα \mathcal{S}_t , $t = 1..p$ για τα οποία οι κόμβοι οι οποίοι ανήκουν στο \mathcal{S}_λ δε συνδέονται με ακμή προς κάποιον κόμβο ο οποίος ανήκει στο \mathcal{S}_κ με $\kappa \neq \lambda$. Σε κάθε \mathcal{S}_t , μπορούμε να βρούμε τον πιο σημαντικό κόμβο εντός του \mathcal{S}_t με βάση την έννοια του eigenvector centrality ([28]) μέσω λίγων επαναλήψεων του αλγόριθμου power iteration.

Η έννοια του eigenvector centrality δείχνει το κατά πόσο οι γείτονες ενός κόμβου i είναι πολλοί σε αριθμό (έννοια γνωστή και σαν degree centrality) αλλά και οι ίδιοι είναι συνδεδεμένοι με πολλούς κόμβους. Είναι γνωστό (λόγω του περιορισμού ότι τα στοιχεία των ιδιοδιανυσμάτων θα είναι θετικά και του σχετικού θεωρήματος Perron–Frobenius) ότι ο προσδιορισμός αυτής της ιδιότητας για έναν κόμβο εξαρτάται από τις μεγαλύτερες ιδιοτιμές (κατ' απόλυτη τιμή) του πίνακα γειτνίασης του γράφου. Ταυτόχρονα, ο αλγόριθμος power iteration αποτελείται από 2 βήματα:

1. choose b_0 (αρχική προσέγγιση του κυρίαρχου ιδιοδιανύσματος, συνήθως με άσους)
2. repeat $b_{k+1} \leftarrow \frac{\mathcal{M}b_k}{\|\mathcal{M}b_k\|}$ until convergence

και εντοπίζει την ιδιοτιμή ενός πίνακα \mathcal{M} με τη μεγαλύτερη απόλυτη τιμή καθώς και το αντίστοιχο ιδιοδιάνυσμα. Αν και η σύγκλιση του αλγορίθμου εν γένει είναι αργή, αν εφαρμοστεί στον πίνακα γειτνίασης \mathcal{M}_t (ο οποίος θα είναι συνήθως μικρής διάστασης) τότε θα μπορέσουμε να προσδιορίσουμε επαρκώς γρήγορα το eigenvector centrality όλων των κόμβων στα αντίστοιχα σύνολα \mathcal{S}_t στα οποία εκείνοι ανήκουν. Για παράδειγμα, έστω η ακόλουθη δομή:



(α) \mathcal{S}_t

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}$$

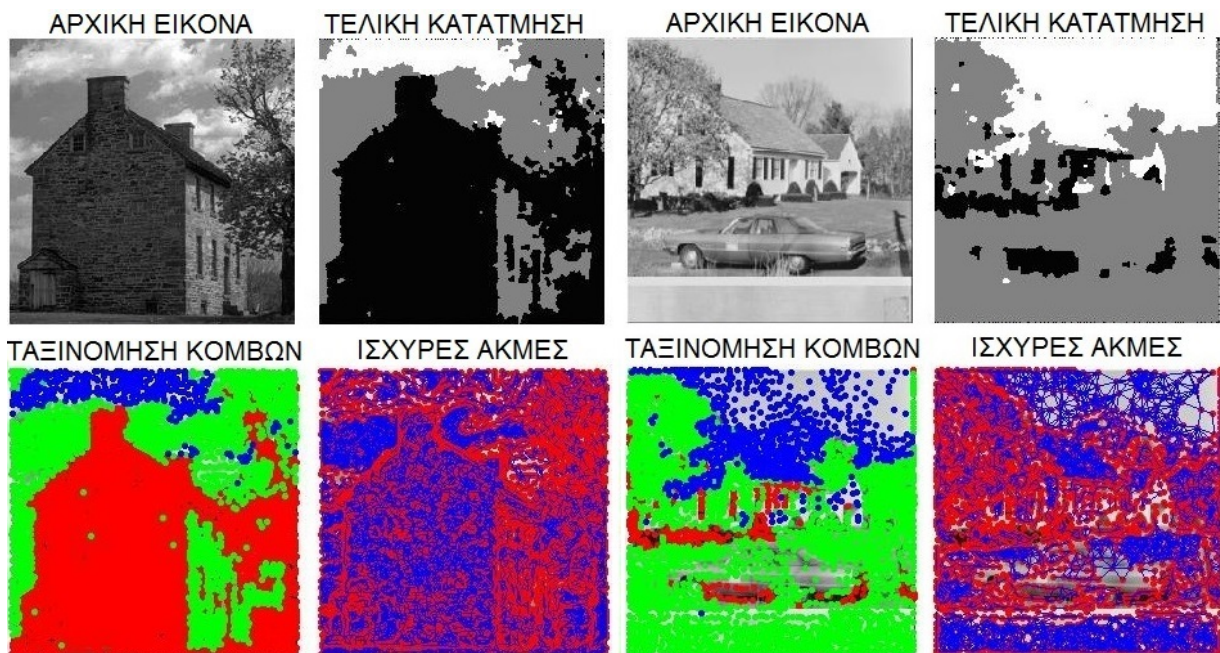
(β) πίνακας γειτνίασης \mathcal{M}_t

Σχήμα 2.8: Πιθανό \mathcal{S}_t και ο αντίστοιχος πίνακας γειτνίασης.

Τα αντίστοιχα centralities των κόμβων προκύπτουν:

$$C = (1.3918 \quad 2.3722 \quad 1.2341 \quad 1.2341 \quad 1.3918 \quad 1.74141)^T$$

Δηλαδή ο κυρίαρχος κόμβος είναι ο 2. Στη συνέχεια, ορίζουμε το πόσα επίπεδα levels θέλουμε να αναδείξουμε στην τελική κατάτμηση και διαλέγουμε για κάθε κυρίαρχο κόμβο την κοντινότερη τιμή που του αντιστοιχεί. Για παράδειγμα, για 5 επίπεδα στο διάστημα $[0, 255]$ θα έχουμε τις πιθανές τιμές $[0 \ 63.75 \ 127.5 \ 191.25 \ 255]$. Αν υπολογίσουμε όλους τους p κυρίαρχους κόμβους των \mathcal{S}_t , τότε κάθε κόμβος ο οποίος $\in \mathcal{S}_t$ αποκτά τη μέση τιμή της φωτεινότητας του αντίστοιχου κυρίαρχου. Σε τελικό στάδιο, τα pixel των περιοχών στις οποίες αρχικά είχαμε τοποθετήσει τους κόμβους λαμβάνουν όλα τη φωτεινότητα του κόμβου που τους αντιστοιχεί, ενώ τα pixel τα οποία βρίσκονται πάνω στα watershed lines παίρνουν την τιμή κάποιας εκ των περιοχών την οποία περιβάλλουν.



Σχήμα 2.9: Αποτέλεσμα κατάτμησης καλής ποιότητας (house) και μέτριας (car + house + yard).

2.3.2 Επέκταση σε 3 διάστασεις - Σχόλια

Ως τώρα περιγράφηκε η περίπτωση $N = 2$, όπου το κατώφλι κινείται πάνω σε μια ευθεία και οι ισχυρές ακμές κινούνται μέσα στο τρίγωνο AOB. Επεκτείνοντας την ιδέα αυτή, ενσωματώνουμε και το feature της διασποράς (sample variance), δηλαδή του πόσο απέχουν οι τιμές φωτεινότητας κάθε περιοχής από τη μέση τιμή (την οποία φέρει ο κόμβος αυτής). Ορίζοντας εκ νέου τις διαφορές μεταξύ αυτών των διασπορών μεταξύ 2 κόμβων αυξάνουμε τη διάσταση του χώρου από 2 σε 3, με τις εξισώσεις να γίνονται:

$$x = x_s, y = y_s, z_s = -\frac{\Gamma}{A}x_s - \frac{\Gamma}{B}y_s + \Gamma$$

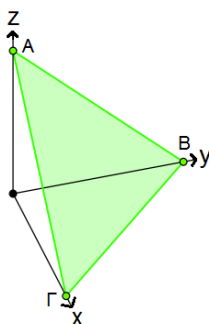
$$edge = \begin{cases} strong & \text{αν } z < z_s \\ weak & \text{αλλιώς} \end{cases}$$

Επομένως, έχουμε τα δύο σημεία (x, y, z) (x_s, y_s, z_s) στον χώρο 3 διαστάσεων, το κατώφλι κινείται πάνω στο τρίγωνο ABΓ και το (x, y, z) , όταν έχουμε ισχυρή ακμή, περιλαμβάνεται στον όγκο που σχηματίζεται. Μια ενδιαφέρουσα ιδιότητα, η οποία επαληθεύτηκε και στην υλοποίηση, είναι ότι η επιλογή των αντιστοιχίσεων A, B, Γ σε απόσταση, φωτεινότητα, διασπορά είναι αυθαίρετη και ισοδύναμη (προκύπτουν δηλαδή τα ίδια υποσύνολα ακμών) για όλες τις περιπτώσεις (προφανώς και για 2 διαστάσεις ισχύει το ίδιο). Ακόμα, το προηγούμενο σετ εξισώσεων θα μπορούσε να γραφτεί και ως (έστω η ίδια αντιστοίχιση των A, B, Γ με πριν):

$$x = x_s, z = z_s, y_s = -\frac{B}{A}x_s - \frac{B}{\Gamma}z_s + B$$

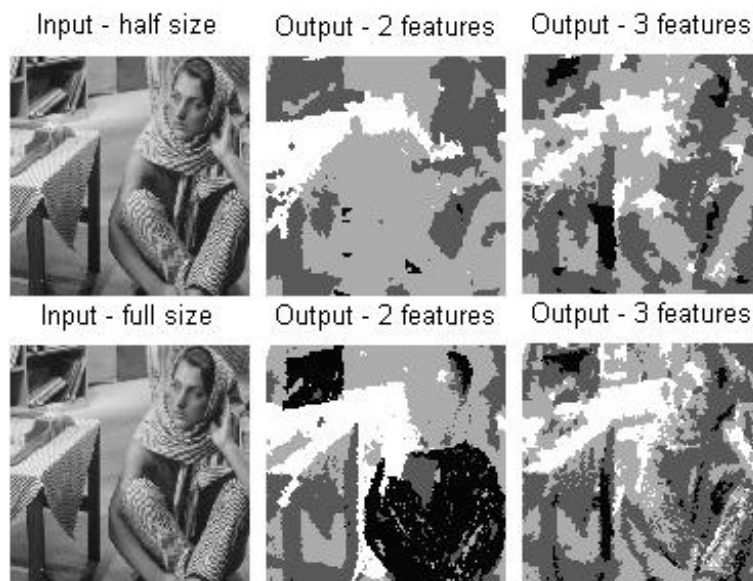
ή ως

$$y = x_s, z = z_s, x_s = -\frac{A}{B}y_s - \frac{A}{\Gamma}z_s + A$$



Σχήμα 2.10: Ο χώρος των ισχυρών ακμών σε 3 διαστάσεις.

Ακολουθούν τα αντίστοιχα αποτελέσματα για 5 και 10 επίπεδα νοητών κόμβων έπειτα από εφαρμογή των παραπάνω σε πραγματικές εικόνες. Σημαντικές παρατηρήσεις προκύπτουν αν αντί των αρχικών εικόνων χρησιμοποιηθούν resized εκδοχές αυτών. Ως προς το resize της εικόνας, υπάρχει μία πολύ έντονη αλλαγή όταν τη μειώσουμε στο μισό. Από τη μία, ο χρόνος εκτέλεσης μειώνεται αλλά και η τελική ποιότητα της κατάτμησης. Ως προς τη διάσταση του επιλεγμένου χώρου (2 ή 3) στην περίπτωση του χώρου 2 διαστάσεων έχουμε χειρότερο αποτέλεσμα από ότι για 3 διαστάσεις, πράγμα που οφείλεται στο ότι επιβίωσαν ακμές όχι τόσο αντιπροσωπευτικές και η εικόνα θλόωσε.



Σχήμα 2.11: Η αρχική εικόνα και η resized εκδοχή αυτής για 5 επίπεδα νοητών κόμβων, woman.



Σχήμα 2.12: Η αρχική εικόνα και η resized εκδοχή αυτής για 10 επίπεδα νοητών κόμβων, woman.

2.4 Συμπεράσματα κεφαλαίου 2 - Ανακεφαλαίωση

Στο κεφάλαιο 2, έγινε μια πρώτη προσπάθεια ανάλυσης της δομής ενός γράφου μέσω της χρήσης μιας συνάρτησης ομοιότητας μεταξύ των κόμβων η οποία αποτιμά τους συσχετισμούς μεταξύ κόμβων. Με τη βοήθεια εφαρμογών πάνω σε πραγματικές εικόνες τις οποίες αναπαριστούμε όχι με pixel αλλά με κόμβους, έγινε αντιληπτή η ευαισθησία του τελικού αποτελέσματος της κατάτμησης από την κύρια παράμετρο της συνάρτησης ομοιότητας, το εύρος σ .

Στη συνέχεια, μελετήθηκαν 2 μέθοδοι (LNP, GLNP) οι οποίες προσπαθούν μέσω κατάλληλα ορισμένων μαθηματικών ελαχιστοποιήσεων να προσδιορίσουν τα βάρη μεταξύ των κόμβων. Η μέθοδος LNP δίνει καλά αποτελέσματα, αλλά εξαρτάται ισχυρά από την παραμέτρο k του k nn γράφου. Αντίστοιχα, η γενίκευση αυτής (GLNP μέθοδος) αναιρεί αυτήν την ευαισθησία, αλλά παραμένει εξαρτώμενη από τη γνώση των clusters στα οποία διαχωρίζουμε τα data points. Η μέθοδος GLNP θα μας απασχολήσει εκ νέου στο κεφάλαιο 4 με μια ισοδύναμη μορφή της μαθηματικής ελαχιστοποίησης που εισάγει.

Τέλος, πραγματοποιήθηκε μια καινούρια προσπάθεια προσδιορισμού της παραμέτρου του εύρους σ με αυτόματο τρόπο με σκοπό την εφαρμογή σε κάποιες εικόνες. Τα αποτελέσματα ήταν ευνοϊκά σε απλές εικόνες με αρκετά ομογενείς περιοχές όχι όμως ικανοποιητικά σε πιο πολύπλοκες εικόνες με αρκετά αντικείμενα και πιο σύνθετο περιβάλλον. Επομένως, καταλαβαίνουμε ότι χρειάζεται να εισάγουμε πιο ισχυρές ιδέες και εργαλεία για να αντιμετωπίσουμε την αντίχνευση αντικειμένων πάνω σε γράφους.

Κεφάλαιο 3

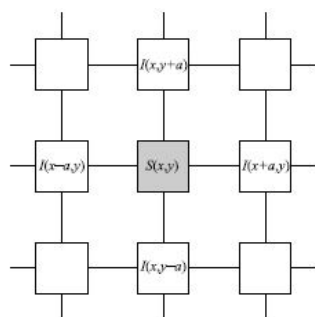
Ελαχιστοποίηση Συναρτησιακών επί του Γράφου - Diffusion

3.1 Μοντέλο επιδημιολογικής εξάπλωσης SIR

3.1.1 Θεωρητικά στοιχεία του μοντέλου

Στη μαθηματική μοντελοποίηση επιδημιολογικών φαινομένων είναι ιδιαίτερα γνωστό το SIR μοντέλο ([12]). Πιο συγκεκριμένα, θεωρούμε 3 ομάδες ανθρώπων: τους susceptible (S), τους infected (I) και τους recovered (R). Ακόμα, θεωρούμε ότι όλοι οι άνθρωποι είναι ακίνητοι και ότι ο συνολικός πληθυσμός N παραμένει αμετάβλητος με το χρόνο ($S + I + R = N$). Ακόμα, θεωρούμε τις συναρτήσεις $S(x, y, t)$, $I(x, y, t)$, $R(x, y, t)$ οι οποίες αν διαιρεθούν με το συνολικό πληθυσμό, αντιστοιχούν στην πιθανότητα του ανθρώπου ο οποίος βρίσκεται στη θέση $v = (x \ y)^T$ να βρίσκεται τη χρονική στιγμή t στην αντίστοιχη κατάσταση (S ή I ή R). Προφανώς, ισχύει ότι $S(x, y, t) + I(x, y, t) + R(x, y, t) = 1 \ \forall t$.

Στη συνέχεια, υποθέτουμε ότι κάθε άνθρωπος, ο οποίος παρίσταται με έναν κόμβο στη θέση v μεταβάλλει την κατάσταση του με βάση την εξής ακολουθία $S \rightarrow I \rightarrow R$, δηλαδή από τη στιγμή που μολυνθεί θα επανακάμψει κάποια χρονική στιγμή αλλά έκτοτε δε γίνεται να μολυνθεί εκ νέου. Η δομή της μεθόδου στηρίζεται στο παρακάτω πλέγμα 4 γειτόνων το οποίο είναι τετραγωνικό διάστασης α :



Σχήμα 3.1: Κανονικό πλέγμα - regular grid.

Η βασική ιδέα του μοντέλου στηρίζεται στο γεγονός ότι κάθε κεντρικός κόμβος στη θέση $v = (x \ y)^T$ δέχεται μια επίδραση σε κάθε χρονική στιγμή t από τους άμεσα γείτο-

νες του πάνω στο καθορισμένο πλέγμα. Η επίδραση αυτή ερμηνεύεται ως η μεταβολή κάθε μιας εκ των 3 συναρτήσεων $S(x, y, t)$, $I(x, y, t)$, $R(x, y, t)$ και είναι ισόποση για κάθε γείτονα. Επομένως, αν ορίσουμε σαν $\Delta I(x, y)$ τη μεταβολή της συνάρτησης $I(x, y, t)$ σε κάθε έναν κεντρικό κόμβο, αναπαριστούμε το μοντέλο με τις ακόλουθες εξισώσεις (αντίστοιχες εξισώσεις προκύπτουν και για τα άλλα δύο μεγέθη, εμείς όμως θα εστιάσουμε μόνο στην ποσότητα $I(x, y, t)$):

$$\Delta I(x, y) = I(x, y, t + 1) - I(x, y, t) \quad (3.1)$$

$$\Delta I(x, y) = \frac{k}{4} S(x, y) (I(x + a, y) + I(x - a, y) + I(x, y + a) + I(x, y - a)) \Delta t \quad (3.2)$$

Σύμφωνα με την (3.2), η πιθανότητα να μολυνθεί ο κόμβος στη θέση (x, y) αυξάνεται ανά χρονικό βήμα Δt κατά μια ποσότητα η οποία είναι ο μέσος όρος των πιθανοτήτων μόλυνσης όλων των άμεσα γειτόνων του πολλαπλασιασμένη με μια κοινή σταθερά k η οποία εκφράζει την ισόποση επίδραση κάθε γειτονικού κόμβου. Επίσης, πολλαπλασιάζουμε και με την πιθανότητα μόλυνσης $S(x, y)$ του κόμβου. Αναπτύσσοντας κατά Taylor κάθε έναν από τους 4 όρους της (3.2) περί το (x, y) έχουμε:

$$I(x + a, y) \cong I(x, y) + aI_x(x, y) + \frac{1}{2}a^2I_{xx}(x, y) \quad (3.3)$$

$$I(x - a, y) \cong I(x, y) - aI_x(x, y) + \frac{1}{2}a^2I_{xx}(x, y) \quad (3.4)$$

$$I(x, y + a) \cong I(x, y) + aI_y(x, y) + \frac{1}{2}a^2I_{yy}(x, y) \quad (3.5)$$

$$I(x, y - a) \cong I(x, y) - aI_y(x, y) + \frac{1}{2}a^2I_{yy}(x, y) \quad (3.6)$$

οπότε προσθέτοντας τις (3.3) έως (3.6) κατά μέλη και αντικαθιστώντας στην (3.2) προκύπτει ότι

$$\Delta I(x, y) = \frac{k}{4} S(x, y) (4I(x, y) + a^2I_{xx}(x, y) + a^2I_{yy}(x, y)) \Delta t \quad (3.7)$$

$$\Rightarrow \Delta I(x, y) = \frac{k}{4} S(x, y) (4I(x, y) + a^2\nabla^2 I(x, y)) \Delta t \quad (3.8)$$

Η τελική σχέση (3.8) στην οποία καταλήξαμε αποτελεί και ένα από τα βασικά εργαλεία μαθηματικής ανάλυσης μιας επιδημιολογικής εξάπλωσης υπό αυτές τις συνθήκες.

3.1.2 Γενίκευση του SIR με σκοπό την εφαρμογή σε αντικείμενα πάνω σε γράφους

Αν κανείς αντικαταστήσει τις 2ες παραγώγους της σχέσης (3.6) με τις διακριτές προσεγγίσεις

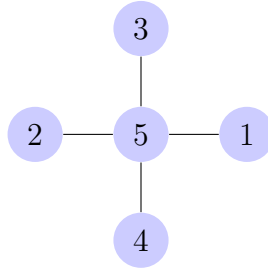
$$I_{xx}(x, y) = \frac{I(x + a, y) - 2I(x, y) + I(x - a, y)}{a^2}$$

$$I_{yy}(x, y) = \frac{I(x, y + a) - 2I(x, y) + I(x, y - a)}{a^2}$$

οδηγείται εκ νέου στην αρχική σχέση (3.2). Επομένως, χρειάζεται να αναμορφώσουμε τις παραπάνω μαθηματικές εξισώσεις με τελικό στόχο την ανίχνευση αντικειμένων πάνω σε γράφους. Αρχικά, θεωρούμε κανονικοποιημένη μονάδα χρόνου ($\Delta t = 1$) καθώς και ότι δεν υπάρχουν άνθρωποι της κατάστασης R, δηλαδή η μόλυνση αποτελεί την τελική κατάσταση κάθε ανθρώπου. Επίσης, θεωρούμε πως όλοι οι κόμβοι είναι με βεβαιότητα στην κατάσταση S ($S(x, y, t) = 1 \forall x, y, t$), δηλαδή κάθε κόμβος είναι με τον ίδιο τρόπο επιρρεπής.

Ακόμα, παρατηρούμε ότι ο όρος $\frac{k}{4}$ στην αρχική σχέση (3.2) θεωρεί ότι κάθε γειτονικός κόμβος συνδράμει με τον ίδιο τρόπο στη μόλυνση του κεντρικού. Κάτι τέτοιο δεν καταφέρνει να αναδείξει την ουσιαστική σχέση μεταξύ των κόμβων της γειτονιάς. Σε ένα πραγματικό παράδειγμα, αν θεωρήσουμε ότι οι κόμβοι αντιπροσωπεύουν ανθρώπους και οι γειτονικοί κόμβοι ενός κεντρικού αντιστοιχίζονται στις φίλιες του ανθρώπου που εξετάζουμε κάθε φορά, οι άνθρωποι με τους οποίους είναι περισσότερο φίλος είναι περισσότερο πιθανό να του μεταδώσουν κάποια ασθένεια καθώς τον συναντάνε πιο συχνά ή μοιράζονται καθημερινές ασχολίες ή αντικείμενα μαζί του. Για το λόγο αυτό, χρησιμοποιούμε τον όρο $\frac{w_{ij}}{N_i}$ (με N_i τον αριθμό των γειτόνων του i κόμβου) ώστε ανάλογα με την ομοιότητα των 2 κόμβων (όρος w_{ij}) να παίρνουμε και ανάλογη συμβολή στο ΔI . Όταν δύο κόμβοι i, j είναι ισχυρά συνδεδεμένοι ($w_{i,j}$ μεγάλο) τότε η επίδραση του i στον j και αντίστροφα είναι μεγαλύτερη. Σε ένα πλέγμα 4 γειτόνων ($N_i = 4 \forall i$) και ονομάζοντας με τον δείκτη 5 τον κεντρικό κόμβο προκύπτει:

$$\Delta I(x, y) = \frac{1}{4}(w_{15}I(x + a, y) + w_{25}I(x - a, y) + w_{35}I(x, y + a) + w_{45}I(x, y - a)) \quad (3.9)$$

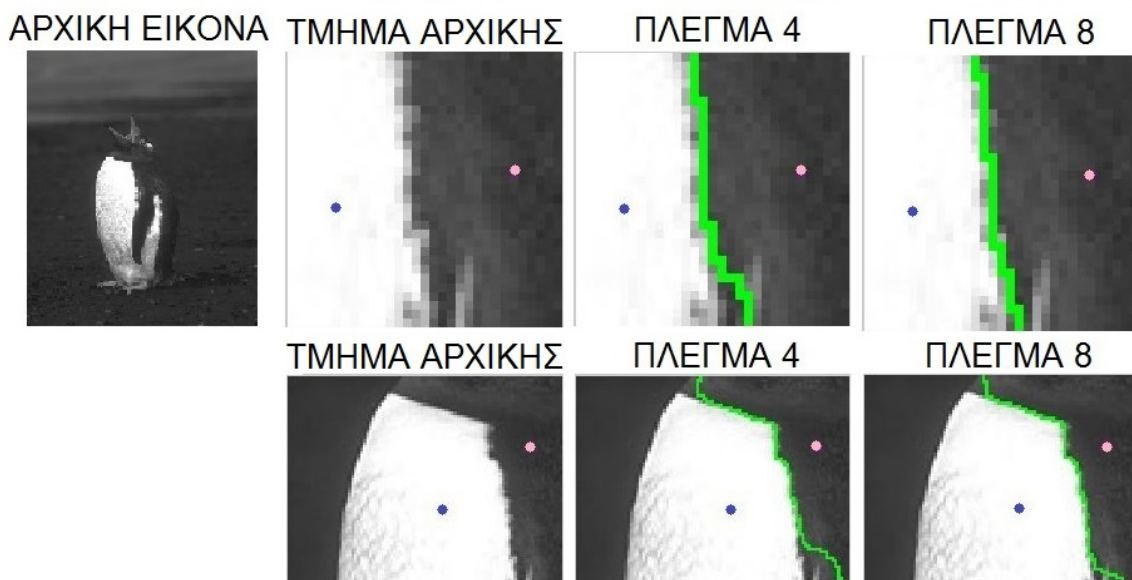


Με τον τρόπο αυτό προκύπτει μια εξίσωση της μορφής: $I_{i,t+1} = I_{i,t} + \Delta I_i$ για κάθε κόμβο i στην εκάστοτε θέση (x, y) . Επαναλαμβάνουμε τη διαδικασία για όλους τους κόμβους και κάνουμε stack τις τιμές σε ένα συνολικό διάνυσμα I_t . Το διάνυσμα αυτό περιέχει σε κάθε timestep την πιθανότητα που υπάρχει σε όλους τους κόμβους της εικόνας να εξαπλωθεί η ασθένεια. Ακόμα, αντικαθιστούμε τον όρο N_i με $\sum_{j \sim i} w_{ij}$.

Για εφαρμογές κατάτμησης εικόνας, απαιτείται η κατάλληλη αναπροσαρμογή της παραπάνω ανάλυσης. Προς τούτο, θεωρούμε ότι γίνεται μετάδοση μέσω γειτονικών κόμβων κάποιας ασθένειας, η οποία περιγράφει ένα αντίστοιχο label για την τελική κατάτμηση (για παράδειγμα για 2 label - foreground και background θα υπάρχουν 2 ασθένειες προς εξαπλώση). Ανά timestep αποφασίζουμε το label (binary 1 ή 0 στο παράδειγμα των 2 ασθενειών) για κάθε κόμβο επιλέγοντας την ασθένεια - label με τη μεγαλύτερη πιθανότητα κάθε φορά, δηλαδή με βάση τον κανόνα: $label_{i,t} = \arg \max_x I_{x,i,t}$. Προφανώς, χρειάζεται να

υπάρχει και κάποιο σημείο στο οποίο εκείνη 'ξέσπασε' (δηλαδή ένα σημείο αρχικοποιημένο στο οποίο έχουμε πιθανότητα 1 ότι η ασθένεια είναι κυρίαρχη στο σημείο αυτό). Επομένως, ορίζουμε manually 1 τουλάχιστον τέτοιο σημείο (σημείο σπόρο) για κάθε ασθένεια που θεωρούμε ότι θα ξεσπάσει στους κόμβους αυτής της εικόνας.

Στο πειραματικό κομμάτι, κάνουμε truncate ένα μικρότερο κομμάτι της αρχικής εικόνας (προς απλοποίηση του προβλήματος) και εκτελούμε τα αντίστοιχα πειράματα για 100 επαναλήψεις και τιμή $\sigma = 90$ (τιμή η οποία συχνά εμφανίζεται στη βιβλιογραφία κυρίως σαν παράμετρος στον αλγόριθμο Τυχαίου Περιπατητή - [7]) για την απόδοση βαρών στις ακμές. Όπως παρατηρούμε παρακάτω, τα αποτελέσματα βελτιώνονται όταν πάμε από πλέγμα 4 γειτόνων σε πλέγμα 8 γειτόνων. Το αποτέλεσμα αυτό είναι λογικό, αν σκεφτεί κανείς ότι η γειτονιά των 8 pixels επιτρέπει πιο ολοκληρωμένες σχέσεις μεταξύ των κόμβων της εικόνας, οπότε και καλύτερα αποτελέσματα κατάτμησης. Παρόλ' αυτά, αν χρησιμοποιήσουμε ένα ευρύτερο κομμάτι της εικόνας, τα αποτελέσματα χειροτερεύουν και για τις 2 επιλογές γειτονιάς.



Σχήμα 3.2: Εξάπλωση ασθένειας για 100 επαναλήψεις σε μικρό/ευρύτερο κομμάτι της εικόνας, penguin.

3.1.3 Καταλήγοντας στην επίλυση της εξίσωσης θερμοτή- τας

Με βάση τα πειραματικά αποτελέσματα της προηγούμενης ενότητας, παρατηρήθηκε ότι η επίλυση της σχέσης (3.9) δε δίνει ικανοποιητικά αποτελέσματα. Για να αναδείξει κανείς τα αίτια αυτής της παρατήρησης, χρειάζεται να κατανοήσει τη μόνιμη κατάσταση η οποία προκύπτει. Στη μόνιμη κατάσταση θα ισχύει $I_{i,t+1} = I_{i,t} = c$ όπου c μια σταθερά, δηλαδή έχουμε κατασταλάξει για τις πιθανότητες εμφάνισης των ασθενειών. Άρα:

$$\text{αν } I_{i,t+1} = I_{i,t} + \Delta I_i \quad I_{i,t+1} = I_{i,t} \quad \Delta I_i = 0 \Rightarrow \sum_{j \sim i} \frac{w_{ij}}{\sum_{j \sim i} w_{ij}} I_{j,t} = 0, \text{ πράγμα αδύνατο}$$

Επομένως, θεωρούμε ότι η τιμή $I_{i,t+1}$ ισούται με $\sum_{j \sim i} \frac{w_{ij}}{\sum_{j \sim i} w_{ij}} I_{j,t}$ μόνο όταν οι τιμές

αυτές των γειτόνων είναι αμετάβλητες, δηλαδή έχουμε φτάσει στη μόνιμη κατάσταση. Η ιδιότητα αυτή για μια συνάρτηση καλείται harmonic property ([4]). Επομένως, ορίζουμε:

$$\Delta I_i = \sum_{j \sim i} w_{ij}(I_{j,t} - I_{i,t}) = \sum_{j \sim i} w_{ij}I_{j,t} - I_{i,t} \sum_{j \sim i} w_{ij} = \sum_{j \sim i} w_{ij}I_{j,t} - I_{i,t}d_i$$

$$\text{άρα } I_{i,t+1} = I_{i,t} + \sum_{j \sim i} w_{ij}I_{j,t} - I_{i,t}d_i$$

Για το υπόλοιπο της μαθηματικής ανάλυσης, απαιτούνται οι εξής μαθηματικοί ορισμοί:

$$U_N = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}, D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_N \end{pmatrix}, d_i = \sum_{j \sim i} w_{ij}$$

$$L = [l^{ij}] = D - W \quad (3.10)$$

$$L_n = [l_n^{ij}] = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}} \quad (3.11)$$

$$l^{ij} = \begin{cases} d_i & \text{αν } i = j \\ -w_{ij} & \text{αν } j \sim i \\ 0 & \text{αλλιώς} \end{cases} \quad l_n^{ij} = \begin{cases} 1 & \text{αν } i = j \\ -\frac{w_{ij}}{\sqrt{d_i d_j}} & \text{αν } j \sim i \\ 0 & \text{αλλιώς} \end{cases}$$

Έστω τώρα v_i^\top η i γραμμή του πίνακα L . Στόχος μας θα είναι ο προσδιορισμός της μόνιμης κατάστασης.

$$v_i^\top I_t = d_i I_{i,t} - \sum_{j \sim i} w_{ij} I_{j,t} = -\Delta I_i \Rightarrow \Delta I_i = -v_i^\top I_t$$

$$\Delta I = \begin{pmatrix} \Delta I_1 \\ \vdots \\ \Delta I_N \end{pmatrix} = - \begin{pmatrix} v_1^\top I_t \\ \vdots \\ v_N^\top I_t \end{pmatrix} = - \begin{pmatrix} v_1^\top \\ \vdots \\ v_N^\top \end{pmatrix} I_t = -L I_t$$

Η PDE με βάση την οποία έγινε η ανάλυση μπορεί να γραφεί και με τη μορφή:

$$I_{i,t+1} = I_{i,t} + \sum_{j \sim i} \partial_j I_{i,t}, \text{ όπου } \partial_j I_{i,t} = w_{i,j}(I_{j,t} - I_{i,t})$$

Παρατηρούμε ότι η επιλογή της κατευθυνόμενης παραγώγου επί μιας ακμής e_{ij} επιλέχθηκε ως η σταθμισμένη διαφορά των τιμών I_i, I_j . Εναλλακτικά, μπορούμε να ορίσουμε μια διαφορετική μόνιμη κατάσταση για έναν κόμβο i , $I_{i,t+1} = \frac{1}{\sqrt{d_i}} \sum_{j \sim i} \frac{w_{ij}}{\sqrt{d_j}} I_{j,t}$ όταν οι τιμές

αυτές των γειτόνων είναι αμετάβλητες. Επομένως, ορίζουμε:

$$I_{i,t+1} = I_{i,t} + \sum_{j \sim i} \partial_j I_{i,t}, \text{ όπου } \partial_j I_{i,t} = \frac{w_{ij}}{\sqrt{d_i}} \left(\frac{I_{j,t}}{\sqrt{d_j}} - \frac{I_{i,t}}{\sqrt{d_i}} \right), \text{ άρα}$$

$$\Delta I_i = \sum_{j \sim i} \frac{w_{i,j}}{\sqrt{d_i}} \left(\frac{I_{j,t}}{\sqrt{d_j}} - \frac{I_{i,t}}{\sqrt{d_i}} \right)$$

Κατ' αναλογία με πριν, έστω v_i^\top η i γραμμή του L_n . Τότε έχουμε:

$$v_i^\top I_t = I_{i,t} - \sum_{j \sim i} \frac{w_{ij}}{\sqrt{d_i d_j}} I_{j,t} = \frac{d_i}{d_i} I_{i,t} - \sum_{j \sim i} \frac{w_{ij}}{\sqrt{d_i d_j}} I_{j,t} \quad (3.12)$$

$$= \sum_{j \sim i} \frac{w_{ij}}{d_i} I_{i,t} - \sum_{j \sim i} \frac{w_{ij}}{\sqrt{d_i d_j}} I_{j,t} = \sum_{j \sim i} \frac{w_{ij}}{\sqrt{d_i}} \left(\frac{I_{i,t}}{\sqrt{d_i}} - \frac{I_{j,t}}{\sqrt{d_j}} \right) \quad (3.13)$$

$$\text{άρα } v_i^\top I_t = -\Delta I_i$$

$$\Delta I = \begin{pmatrix} \Delta I_1 \\ \vdots \\ \Delta I_N \end{pmatrix} = - \begin{pmatrix} v_1^\top I_t \\ \vdots \\ v_N^\top I_t \end{pmatrix} = - \begin{pmatrix} v_1^\top \\ \vdots \\ v_N^\top \end{pmatrix} I_t = -L_n I_t$$

Άρα: $I_{t+1} = I_t - L_n I_t = (U_N - L_n) I_t$. Θέτοντας $R = U_N - L_n$ παίρνουμε: $I_{t+1} = R I_t = R^{t+1} I_0$, όπου I_0 είναι τα label που μας έχουν δοθεί (1 όταν υπάρχει η ασθένεια στο σημείο και 0 αλλιώς για 2 ασθένειες). Στη συνέχεια, κάνουμε την εξής αναδιάταξη και partition των R και I_t : αντιστοιχίζουμε σε έναν πίνακα R_U όλες τις γραμμές του πίνακα R οι οποίες αντιστοιχούν σε κόμβους οι οποίοι είναι unlabeled (οι υπόλοιπες γραμμές αντιστοιχίζονται στον πίνακα R_L) και αντίστοιχα παίρνουμε και το διάνυσμα $I_{U,t}$ για τη χρονική στιγμή t .

$$R = \begin{pmatrix} R_U \\ R_L \end{pmatrix}, I_t = \begin{pmatrix} I_{U,t} \\ I_{L,t} \end{pmatrix}, \text{ με το } I_{L,t} \text{ να ταυτίζεται με το } I_{L,0} \text{ (seeds)} \forall t$$

$$I_{U,t+1} = R_U I_t$$

Επιπρόσθετα, κάνουμε ακόμα την εξής αναδιάταξη και partition του R_U : αντιστοιχίζουμε σε έναν πίνακα R_{UU} όλες τις στήλες του πίνακα R_U οι οποίες αντιστοιχούν σε κόμβους οι οποίοι είναι unlabeled (οι υπόλοιπες στήλες αντιστοιχίζονται στον πίνακα R_{UL}).

$$R_U = \begin{pmatrix} R_{UU} \\ R_{UL} \end{pmatrix}^\top$$

επομένως, για $C = R_{UL} I_{L,0}$ έχουμε:

$$I_{U,t+1} = R_U I_t = R_{UU} I_{U,t} + R_{UL} I_{L,0} \quad (3.14)$$

$$= R_{UU} (R_{UU} I_{U,t-1} + R_{UL} I_{L,0}) + R_{UL} I_{L,0} \quad (3.15)$$

$$= R_{UU}^2 I_{U,t-1} + R_{UU} C + C \quad (3.16)$$

συνεχίζοντας αναδρομικά και για $Z = N - N_{seeds}$, έχουμε:

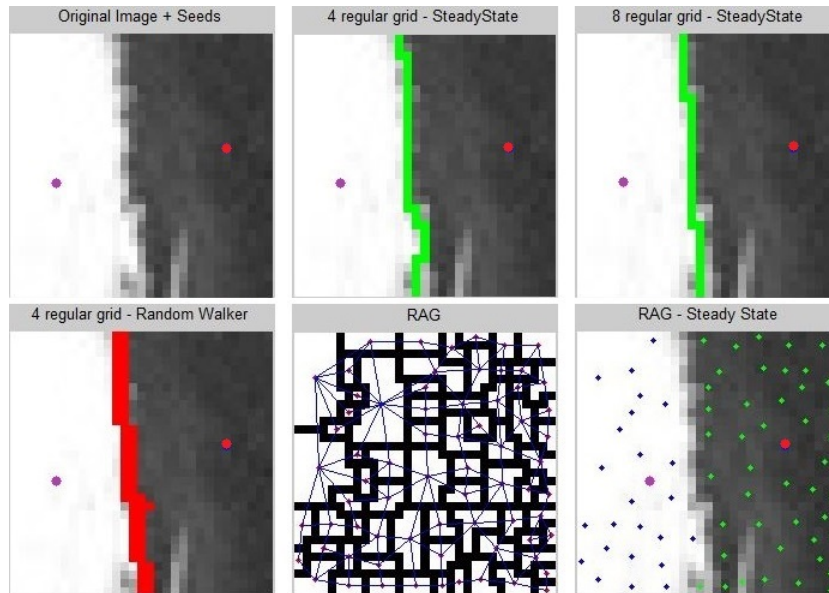
$$I_{U,t+1} = R_{UU}^{t+1} I_{U,0} + \sum_{\kappa=0}^t R_{UU}^{\kappa} C \stackrel{I_{U,0} \stackrel{=}{=} 0}{t \rightarrow \infty} (U_Z - R_{UU})^{-1} C = (U_Z - R_{UU})^{-1} R_{UL} I_{L,0}$$

Επομένως, οδηγηθήκαμε σε έναν παρόμοιο τρόπο ελαχιστοποίησης (βλέπε [18]) του

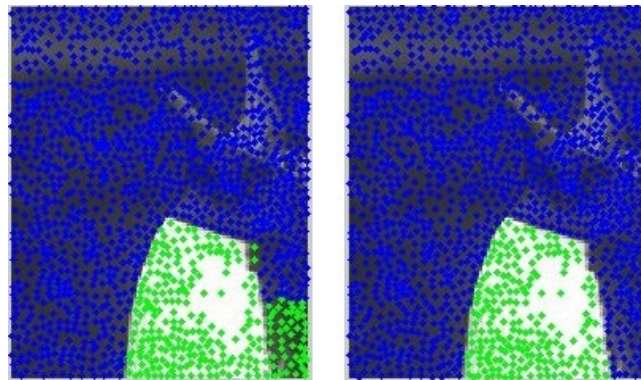
$$J(I) = \frac{1}{2} \sum_{i,j=1}^N w_{i,j} \left(\frac{I_i}{\sqrt{d_i}} - \frac{I_j}{\sqrt{d_j}} \right)^2 = \frac{1}{2} I^T L_n I$$

3.1.4 Η δυναμική εξέλιξη της εξάπλωσης - Αποτελέσματα

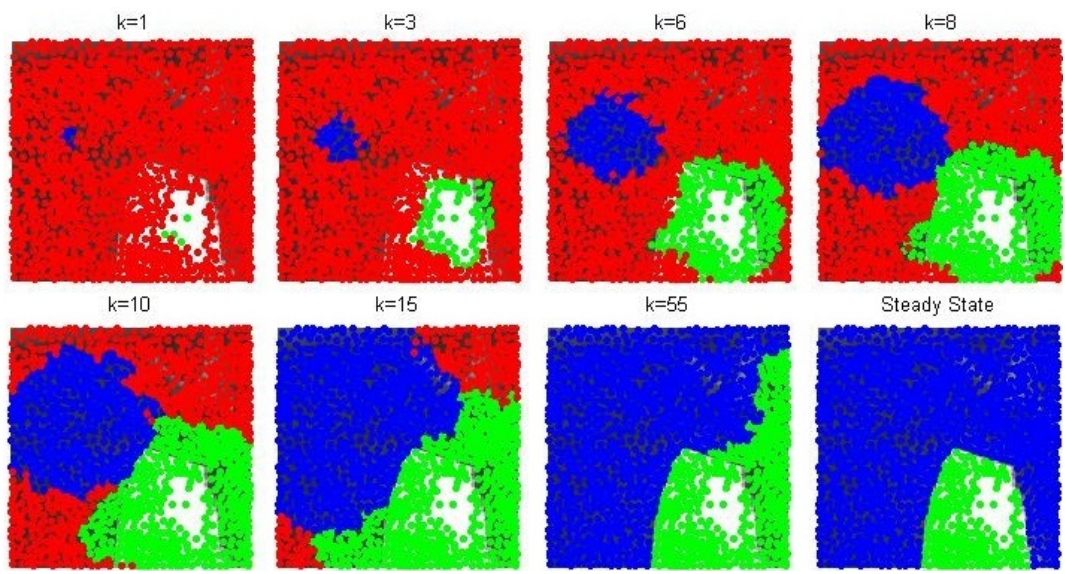
Με βάση την προηγούμενη ανάλυση της εξάπλωσης 2 ασθενειών επί του γράφου \mathcal{G} , παρουσιάζονται αντίστοιχα αποτελέσματα. Σημειώνεται ότι παρουσιάζονται τα αποτελέσματα που αφορούν την ελαχιστοποίηση της κανονικοποιημένης Graph Laplacian L_n , ενώ ο αλγόριθμος Random Walker αφορά τη Graph Laplacian L .



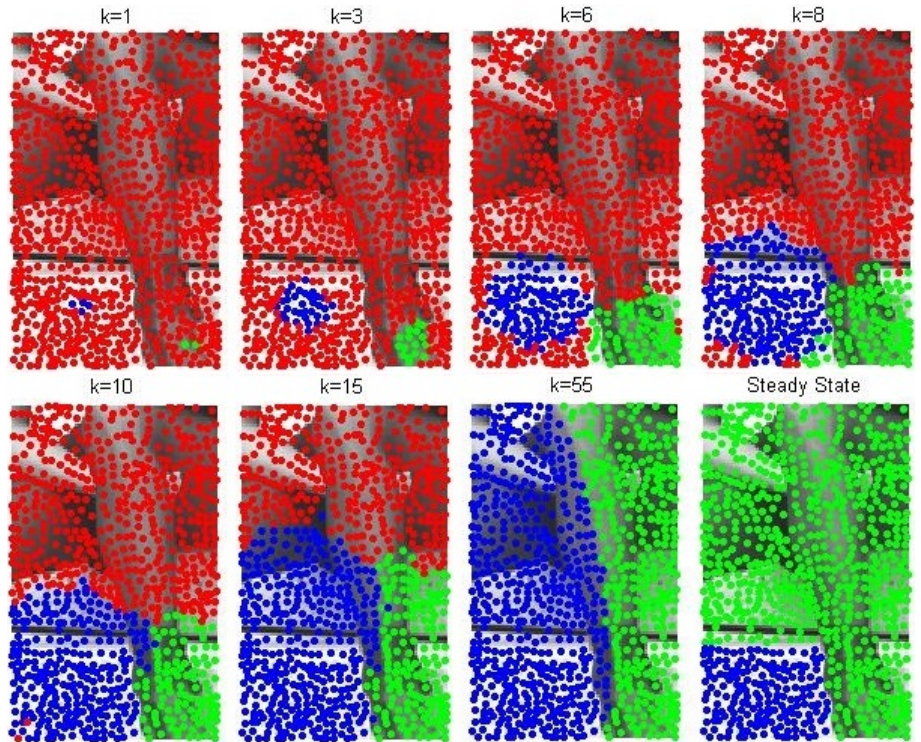
Σχήμα 3.3: Αποτελέσματα μόνιμης κατάστασης με διάφορες μεθόδους, penguin piece 1.



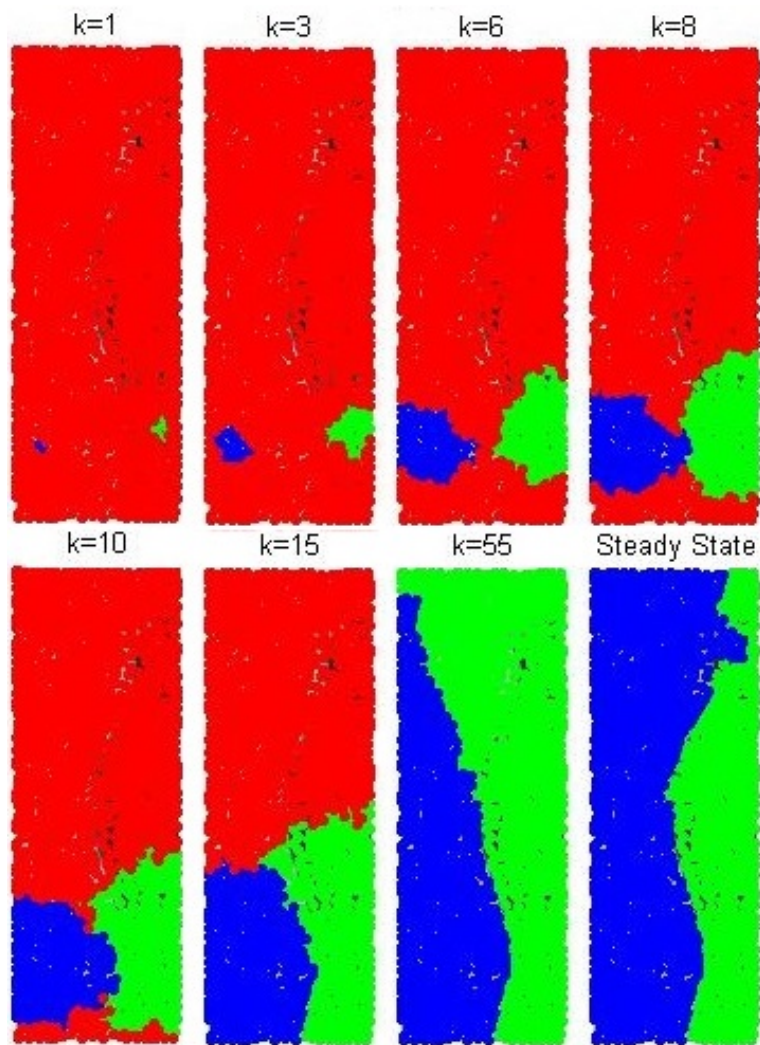
Σχήμα 3.4: Εξάπλωση 2 ασθενειών (500 επαναλήψεις) και μόνιμη κατάσταση, penguin piece 3.



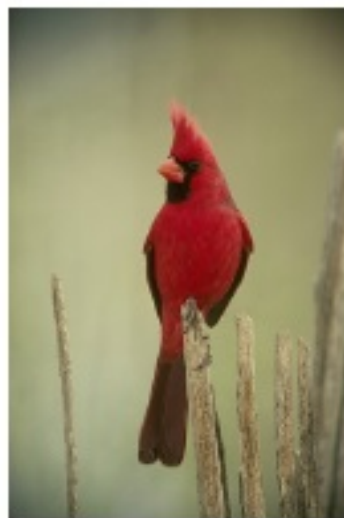
Σχήμα 3.5: Εξάπλωση 2 ασθενειών και μόνιμη κατάσταση, penguin piece 3.



Σχήμα 3.6: Εξάπλωση 2 ασθενειών και μόνιμη κατάσταση, woman + stairs piece.



Σχήμα 3.7: Εξάπλωση 2 ασθενειών και μόνιμη κατάσταση, bird piece.



Σχήμα 3.8: Αρχικές εικόνες, woman + stairs, bird.

3.2 Η εξίσωση της θερμότητας πάνω σε γράφους - label propagation

3.2.1 Ελαχιστοποίηση με βάση τη νόρμα L_2

Η τεχνική του label propagation ([5]) στοχεύει στη διάχυση των label τα οποία είναι διαθέσιμα για ένα μικρό σύνολο data points (labeled) σε εκείνα τα οποία είναι unlabeled. Στηρίζομενοι στην ιδέα ότι data points τα οποία μοιάζουν μεταξύ τους θα πρέπει να αποκτούν το ίδιο label, το label propagation στηρίζεται στην ελαχιστοποίηση του συναρτησιακού

$$J(f) = \frac{1}{2} \sum_{i,j=1}^N w_{i,j} (f_i - f_j)^2$$

όπου f η συνάρτηση απόδοσης ετικέτας στα σημεία. Στη βιβλιογραφία έχουν ακολουθηθεί και παρόμοιες ιδέες, ελαχιστοποιώντας το συναρτησιακό

$$J(f) = \frac{1}{2} \sum_{i,j=1}^N w_{i,j} \left(\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \right)^2$$

αλλά και με άλλες τεχνικές, ([4], [7], [13]). Ακόμα, είναι γνωστό ότι οι παραπάνω ιδέες σχετίζονται άμεσα με την επίλυση της εξίσωσης της θερμότητας ($u_t = \nabla^2 u$) πάνω στο γράφο \mathcal{G} ([4]). Η βασική ιδέα πίσω από τη διάχυση θερμότητας είναι ότι υπάρχει μεταφορά ενέργειας από τις θερμές περιοχές (περιοχές υψηλής ενέργειας) στις πιο ψυχρές (χαμηλότερης ενέργειας). Η αντιστοιχία με το label propagation είναι ότι τα προσημασμένα σημεία είναι σημεία υψηλής ενέργειας και τα unlabeled είναι χαμηλότερης. Συνεπώς, η ενέργεια (απόδοση ετικέτας) διαχέεται πάνω στο γράφο οδηγούμενη από την αρχή αυτή. Επομένως, καταλαβαίνουμε ότι η προηγούμενη μας ανάλυση η οποία ξεκίνησε από το επιδημιολογικό μοντέλο SIR, κατέληξε στην επίλυση της διάχυσης θερμότητας επί του γράφου, ένα πρόβλημα πολύ πιο γνωστό και οικείο σε εφαρμογές Όρασης Υπολογιστών (από τη σκοπιά των diffusion διαδικασιών) και Εκμάθησης Μηχανών (από τη σκοπιά του label propagation).

3.2.2 Ελαχιστοποίηση με βάση τη νόρμα L_1 - Αντιμετώπιση outliers

Στη μελέτη του label propagation ([1], [5]) όπως και στον αλγόριθμο Random Walker ([7]) έχει δοθεί ιδιαίτερη έμφαση στην ελαχιστοποίηση της L_2 νόρμας. Σχετικά πρόσφατα, έχουν αναπτυχθεί ιδέες γύρω από sparse coding οι οποίες στηρίζονται στην ελαχιστοποίηση της L_1 νόρμας, επιδιώκουν δηλαδή λύσεις με πολλά μηδενικά στοιχεία.

Είναι γνωστό ότι η μέθοδος των ελαχίστων τετραγώνων μπορεί να αντιμετωπίσει σε ικανοποιητικό βαθμό την ύπαρξη θορύβου, όταν αυτός μοντελοποιείται μέσω μιας Gaussian κατανομής. Πιο συγκεκριμένα, έστω $\hat{X} \in \mathbf{R}^{m \times n}$ η λαμβανόμενη μέτρηση. Αν θεωρήσει κανείς την ύπαρξη προσθετικού θορύβου $E \in \mathbf{R}^{m \times n}$ με $E \sim \mathcal{N}(0, \sigma^2)$ τότε έχουμε για τις πραγματικές μετρήσεις $X = \hat{X} - E$, με $X \in \mathbf{R}^{m \times n}$. Η υπόθεση ύπαρξης προσθετικού θορύβου είναι ρεαλιστική για πληθώρα εφαρμογών, όπως για την περίπτωση των εικόνων.

Μια ισοδύναμη πιθανοτική ερμηνεία είναι ότι η μεγιστοποίηση της $\log P(\hat{X}|X)$ ταυτίζεται με την ελαχιστοποίηση του αθροίσματος των τετραγωνών των διαφορών μεταξύ των στοιχείων του πίνακα \hat{X} και X . Δυστυχώς, η υπόθεση ότι ο θόρυβος ακολουθεί μια τέτοια κατανομή δεν είναι επαρκής, οπότε και εισάγονται ελαχιστοποιήσεις τύπου L_1 οι οποίες αντιμετωπίζουν καλύτερα το θόρυβο ([8],[9]). Έστω, λοιπόν, το κλασικό πρόβλημα ελαχιστοποίησης

$$\min_f Q(f) = \frac{\lambda}{2} f^\top L f + \frac{1}{2} \|f - y\|^2$$

Θεωρούμε ότι για τον L ισχύει:

$$L = V \Sigma V^\top$$

όπου ο V είναι ένας $n \times n$ πίνακας που περιέχει σαν στήλες τα ιδιοδιανύσματα του L και ο $\Sigma = [\sigma_{ii}]$ ένας $n \times n$ διαγώνιος πίνακας με σ_{ii} τις ιδιοτιμές ($0 \leq \sigma_{11} \leq \dots \leq \sigma_{nn}$). Τότε, έχουμε:

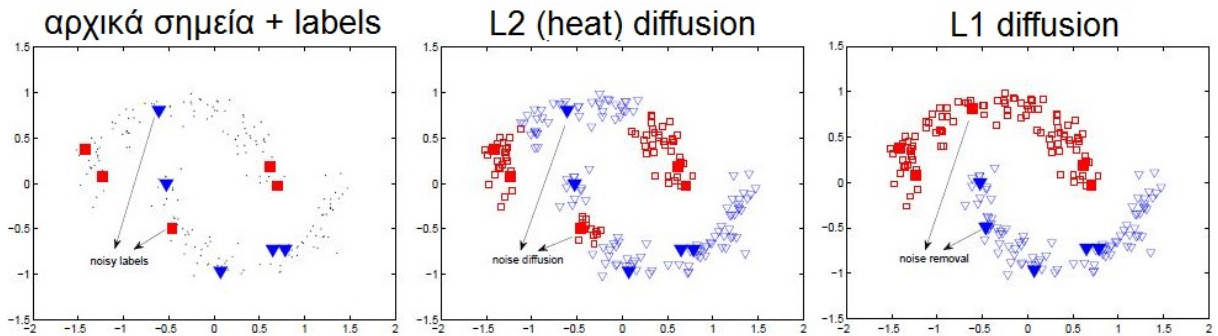
$$L = V \Sigma V^\top = (\Sigma^{\frac{1}{2}} V^\top)^\top \Sigma^{\frac{1}{2}} V^\top = B^\top B, B = \Sigma^{\frac{1}{2}} V^\top$$

Ορίζουμε την ποσότητα $\Omega(f) = \frac{\lambda}{2} f^\top L f$ και $\tilde{\Omega}(f) = \|Bf\|_1$ και αποδεικνύεται ([9]) ότι αν $\tilde{\Omega}(f) \leq 1$ τότε $\Omega(f) \leq \tilde{\Omega}(f)$. Επομένως, μπορούμε να χρησιμοποιήσουμε σα regularizer το $\tilde{\Omega}(f)$ αντί του $\Omega(f)$ και επιπλέον θεωρούμε ότι η λύση f μπορεί να γραφεί ως $V_m \alpha$ όπου V_m είναι ένας $n \times m$ πίνακας ο οποίος περιέχει τα m ιδιοδιανύσματα με τις μικρότερες ιδιοτιμές (δηλαδή τις m πρώτες στήλες του V). Με τον τρόπο αυτό, το πρόβλημα μετατρέπεται στο:

$$\tilde{Q}(f) = \lambda \|Bf\|_1 + \frac{1}{2} \|f - y\|^2 \text{ οπότε για } f = V_m \alpha$$

$$\tilde{Q}(\alpha) = \lambda \left\| \Sigma^{\frac{1}{2}} V^\top V_m \alpha \right\|_1 + \frac{1}{2} \|V_m \alpha - y\|^2 = \lambda \sum_{i=1}^m \Sigma_{ii}^{\frac{1}{2}} |\alpha_i| + \frac{1}{2} \|V_m \alpha - y\|^2$$

Το παραπάνω πρόβλημα επιλύεται με διάφορους αλγόριθμους, όπως ο FISTA ([10]). Χαρακτηριστικό αποτέλεσμα είναι η μείωση της επίδρασης του θορύβου των αρχικών label στο τελικό diffusion:



Σχήμα 3.9: Αρχικά label και τελικό L_2 και L_1 diffusion, από [9].

3.3 Αλγόριθμος Τυχαίου Περιπατητή (Random Walker)

3.3.1 Εισαγωγικά στοιχεία της μεθόδου

Ο αλγόριθμος του Τυχαίου Περιπατητή (Random Walker) ([7]) αποτελεί ένα πολύ διαδεδομένο εργαλείο κατάτμησης εικόνων, με πληθώρα εφαρμογών σε πολλούς τομείς (όπως οι ιατρικές εφαρμογές) με σκοπό τον εντοπισμό περιοχών ενδιαφέροντος. Η μέθοδος αυτή ανήκει στις μεθόδους με επίβλεψη, δηλαδή ο χρήστης ορίζει κάποια σημεία (seeds) πάνω στην εικόνα τα οποία θεωρεί χαρακτηριστικά σημεία των περιοχών που θέλει να ξεχωρίσει. Το τελικό αποτέλεσμα ερμηνεύεται σαν την πιο πιθανή ετικέτα s κάθε κόμβου, δηλαδή την εύρεση εκείνου του σπόρου στον οποίο είναι πιο πιθανή η κατάληξη ενός τυχαίου περιπατητή ο οποίος ξεκινά τη διαδρομή του από κάθε μη προσημασμένο κόμβο.

3.3.2 Μαθηματική διατύπωση της παραδοσιακής μεθόδου Random Walker

Έστω γράφος $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ με κόμβους $v \in \mathcal{V}$ και ακμές $e \in \mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Έστω ακόμη ότι $m = |\mathcal{E}|$ και $n = |\mathcal{V}|$. Κάθε ακμή e η οποία συνδέει δύο κόμβους i και j χαρακτηρίζεται ως e_{ij} . Υποθέτουμε ότι ο γράφος \mathcal{G} είναι συνεκτικός, μη κατευθυνόμενος και με βάρη $w_{ij} > 0$ τα οποία καθορίζουν τη συμπεριφορά του τυχαίου περιπατητή επί της ακμής e_{ij} . Μεγάλες τιμές του w_{ij} ενθαρρύνουν τον περιπατητή να περάσει από αυτήν την ακμή, ενώ μικρές των αποθαρρύνουν. Σημειώνεται ότι αν όλα τα βάρη ήταν μοναδιαία τότε ο περιπατητής θα ήταν τυχαίος με την αυστηρή έννοια του όρου. Ορίζουμε ακόμα το βαθμό του κάθε κόμβου $d_i = \sum_{j \sim i} w_{ij}$ και μοντελοποιούμε την ομοιότητα μεταξύ δύο κόμβων i, j με το

γνωστό Gaussian πυρήνα με το εύρος σ να αποτελεί τη μοναδική ελεύθερη παράμετρο. Η παράμετρος αυτή ορίζεται από το χρήστη και επηρεάζει κυρίως το τελικό αποτέλεσμα της κατάτμησης.

Η απευθείας επίλυση του προβλήματος του Τυχαίου Περιπατητή είναι υπολογιστικά πολύ δύσκολη οπότε έχει προταθεί εναλλακτικά η λύση του Dirichlet προβλήματος του οποίου η λύση είναι γνωστό ότι είναι ταυτόσημη με τις ζητούμενες πιθανότητες μετάβασης των κόμβων στα σημεία σπόρους. Το πρόβλημα αυτό έγκειται στην ελαχιστοποίηση του Dirichlet ολοκληρώματος $D[u] = \frac{1}{2} \int_{\Omega} |\nabla u|^2 d\Omega$ για ένα πεδίο u σε μια περιοχή Ω . Η εξίσωση Laplace είναι η εξίσωση Euler - Lagrange ([26]) για το ολοκλήρωμα Dirichlet οπότε και αναζητάμε λύση της εξίσωσης Laplace δεδομένων των αρχικών συνθηκών, οι οποίες αντιστοιχούν στα σημεία που έχει δώσει ο χρήστης. Με βάση τον παραδοσιακό ορισμό του graph Laplacian, μπορούμε να δούμε ότι $\mathcal{L} = \mathcal{A}^T \mathcal{C} \mathcal{A}$ όπου $\mathcal{A} \in \mathcal{R}^{m \times n}$, $\mathcal{C} \in \mathcal{R}^{m \times m}$ και $\mathcal{L} \in \mathcal{R}^{n \times n}$ με τον \mathcal{C} να περιέχει στη διαγώνιο του τα βάρη των ακμών και τον \mathcal{A} να είναι ο πίνακας πρόσπτωσης του γράφου.

Επομένως, το ολοκλήρωμα *Dirichlet* στη συνδυαστική - διακριτοποιημένη εκδοχή του γράφεται: $\mathcal{D}[x] = \frac{1}{2}(\mathcal{A}x)^\top \mathcal{C}\mathcal{A}x = \frac{1}{2}x^\top \mathcal{A}^\top \mathcal{C}\mathcal{A}x = \frac{1}{2}x^\top \mathcal{L}x = \frac{1}{2} \sum_{j \sim i} w_{ij}(x_i - x_j)^2$

Τα κρίσιμα σημεία του $\mathcal{D}[x]$ είναι σημεία ελαχίστου και μεγίστου. Επειδή, όμως, ο τελεστής \mathcal{L} είναι θετικά ημι-ορισμένος, τα κρίσιμα σημεία ταυτίζονται με τα σημεία ελαχίστου τα οποία ψάχνουμε. Παραγωγίζοντας τον τελεστή $\mathcal{D}[x]$ και ακολουθώντας τη συλλογιστική του section 3.1.3 διαχωρίζουμε τον \mathcal{L} ως

$$\mathcal{L} = \begin{pmatrix} \mathcal{L}_M & \mathcal{B} \\ \mathcal{B} & \mathcal{L}_U \end{pmatrix}$$

όπου οι υποδείκτες M, U δείχνουν το προσημασμένο (marked) ή unmarked κομμάτι του πίνακα \mathcal{L} ή του x . Τελικά, καταλήγει κανείς στη λύση του εξής γραμμικού συστήματος: $\mathcal{L}_U x_U = -\mathcal{B}^\top x_M$. Από το γεγονός ότι οι πιθανότητες κατάληξης ενός unmarked κόμβου στα σημεία σπόρους (οι οποίοι ανήκουν σε K συνολικά ετικέτες) αθροίζονται στη μονάδα, δηλαδή $\sum_i x_i^s = 1$, προκύπτει ότι απαιτείται η λύση $K - 1$ τέτοιων γραμμικών συστημάτων και όχι K .

3.3.3 Σύνδεση με άλλες μεθόδους

Ο αλγόριθμος του Τυχαίου Περιπατητή (Random Walker) συνδέεται στενά και με άλλες ιδέες οι οποίες έχουν μελετηθεί διεξοδικά. Τα random walks συνδέονται στενά με τη θεωρία των ηλεκτρικών κυκλωμάτων ([27]). Συνοπτικά, αναφέρουμε ότι αν θεωρήσει κανείς ρεύματα z , πηγές έντασης f , πηγές τάσης b και τάσεις p , μπορούμε να συνδυάσουμε τους εξής 3 νόμους της θεωρίας κυκλωμάτων $\mathcal{A}^\top z = f$ (Νόμος Ρευμάτων Kirchoff), $\mathcal{C}p = z$ (Νόμος του Ohm) και $p = \mathcal{A}x + b$ (Νόμος Τάσεων Kirchoff) σε μια ενιαία εξίσωση της μορφής

$$\mathcal{A}^\top \mathcal{C}\mathcal{A}x + \mathcal{A}^\top \mathcal{C}b = \mathcal{L}x + \mathcal{A}^\top \mathcal{C}b = f \Rightarrow \mathcal{L}x = f - \mathcal{A}^\top \mathcal{C}b$$

εξίσωση η οποία για $f = 0$ αντιστοιχεί στη βασική εξίσωση του αλγόριθμου.

Η δεύτερη σύνδεση (με το diffusion) θα μας απασχολήσει περισσότερο, καθώς θα μπορούσαμε μέσω αυτής να χρησιμοποιήσουμε μια τροποποίηση του παραδοσιακού αλγόριθμου Τυχαίου Περιπατητή. Υπενθυμίζεται ότι η εξίσωση θερμότητας ($u_t = \nabla^2 u$) συνδέθηκε προηγουμένως με τις έννοιες του diffusion και του label propagation. Αν θεωρήσει κανείς τη μόνιμη κατάσταση ($t \rightarrow \infty$) ή, ισοδύναμα, τον τερματισμό της εξάπλωσης των ασθενειών στο μοντέλο του SIR με τον τρόπο που το χειριστήκαμε, παρατηρεί ότι λύση της εξίσωσης Laplace $\nabla^2 u = 0$ είναι η μόνιμη κατάσταση στην επίλυση του diffusion και ταυτόχρονα η κατάσταση στην οποία όλοι οι κόμβοι έχουν αμετάβλητη πιθανότητα να πάσχουν από κάποια ασθένεια, καθώς οι γείτονες τους, όπως και οι ίδιοι, έχουν σταθερή τέτοια πιθανότητα, ίση με το σταθμισμένο μέσο όρο των αντίστοιχων γειτόνων τους. Η αναλογία αυτών των πιθανοτήτων με τις αντίστοιχες πιθανότητες να καταλήξει ο τυχαίος περιπατητής σε κάποιο σπόρο ετικέτας είναι εμφανής και αναδεικνύει τη σύνδεση αυτή.

Αξίζει να σημειωθεί ότι στην ανάλυση η οποία έγινε στο SIR ώστε να προκύψουν οι αντίστοιχες PDE's, δεν εισάγαμε τον περιορισμό οι πιθανότητες να υπάρχει κάποια ασθένεια

να ανθροίζονται στη μονάδα, απλά εξασφαλίσαμε ότι αυτές θα είναι φραγμένες από τη μονάδα. Επομένως, δε θεωρήσαμε την έννοια της πιθανότητας με την αυστηρή μορφή αυτής.

3.3.4 Επεκτάσεις του Αλγορίθμου Random Walker

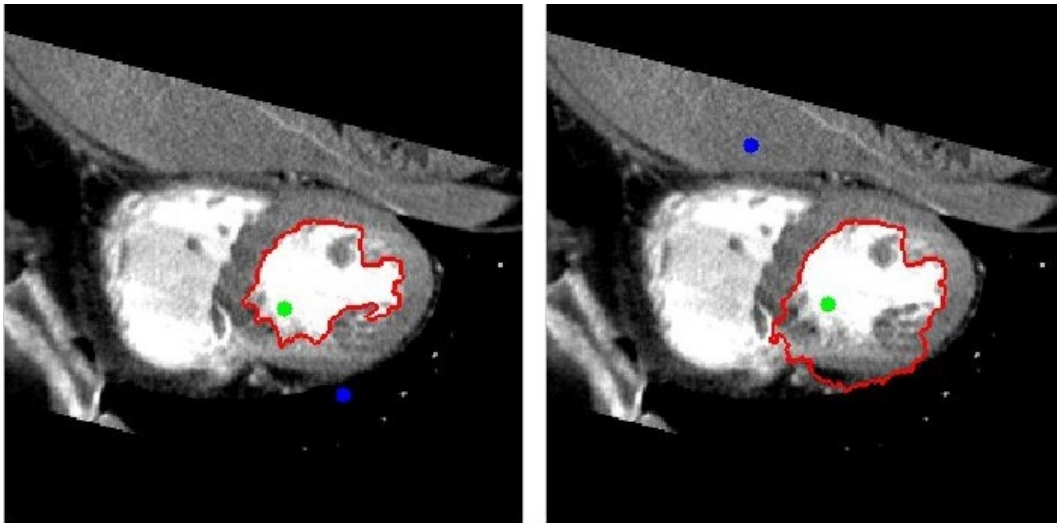
Έχουν προταθεί πολλές επεκτάσεις στον παραδοσιακό αλγόριθμο του Random Walker. Για παράδειγμα, έχει προταθεί η προσθήκη prior models ώστε να ανιχνεύονται τμήματα των αντικειμένων τα οποία δεν είναι συνδεδεμένα σε κάποιο σημείο σπόρο ([23]). Επίσης, στο ([24]) προτείνεται η χρήση filter bank features με στόχο την κατηγοριοποίηση/κατάταξη υφής. Η τροποποίηση του γράφου σε περιοχές οι οποίες προκύπτουν από κάποια υπερκατατιμημένη εκδοχή της εικόνας έχει μελετηθεί στο [17] με σκοπό την πιο γρήγορη επεξεργασία της εικόνας και τη μείωση των κόμβων του γράφου. Τέλος, στο [25] εξετάζεται η επιτάχυνση του αλγόριθμου με τεχνικές precomputation.

3.3.5 Περιορισμοί και πλεονεκτήματα του παραδοσιακού αλγόριθμου Random Walker

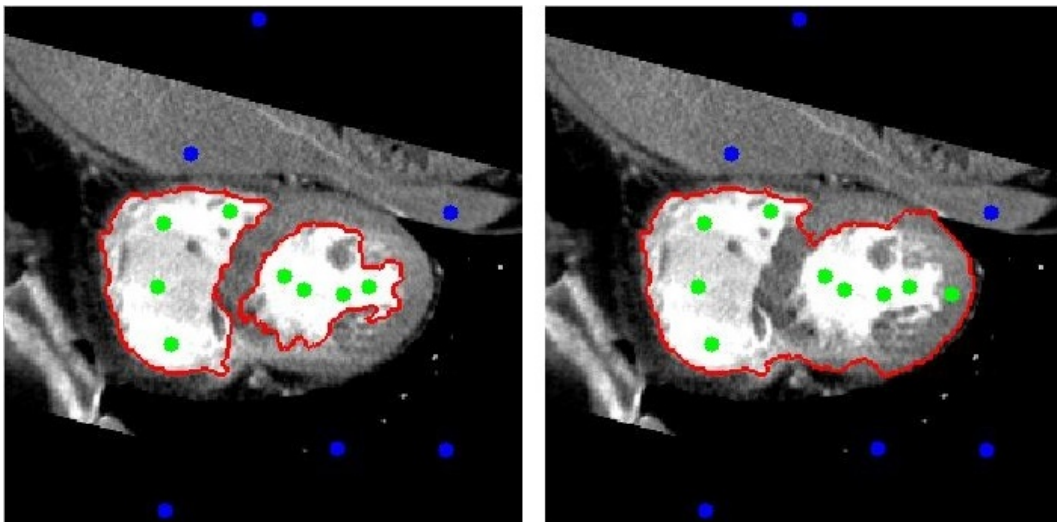
Περιορισμοί του παραδοσιακού Random Walker:

1. Η μέθοδος είναι με επίβλεψη, οπότε ο ανθρώπινος παράγοντας μπορεί να επιδράσει κατασταλτικά. Για παράδειγμα, το extraction προσημασμένων δεδομένων είναι δύσκολη διαδικασία και τις περισσότερες φορές ο αριθμός τους είναι πάρα πολύ μικρότερος από εκείνων τα οποία καλούμαστε να αναλύσουμε. Έτσι, είναι δύσκολη η γενίκευση στο σύνολο των δεδομένων.
2. Η ποιότητα των προσημασμένων δεδομένων είναι όχι μόνο περιορισμένη, αλλά κυρίως μπορεί να είναι χαμηλής ποιότητας. Το στοιχείο αυτό, σε συνδυασμό με την ευαισθησία του αλγορίθμου στα outliers (μιας και χρησιμοποιεί έναν όρο L_2 smoothness), ενδέχεται να μειώσει την αποτελεσματικότητά του.
3. Η παράμετρος σ είναι δύσκολο να προσδιοριστεί με κάποιον αυτόματο τρόπο, όπως είδαμε και στο Κεφάλαιο 2. Έχουν προταθεί κάποιες μέθοδοι τοπικού προσδιορισμού της (ενδεικτικά στο [11]) αλλά δύσκολα μπορούμε να βρούμε κάποιο γενικό κανόνα. Κυριότερος τρόπος προσδιορισμού είναι με τεχνικές cross - validation.
4. Όπως ακριβώς και το label propagation, έτσι και ο αλγόριθμος Τυχαίου Περιπατητή (τον οποίο μπορούμε να σκεφτούμε και σαν τη μόνιμη κατάσταση του label propagation) είναι επιρρεπής στην επιλογή των αρχικών σημείων.

Οι παραπάνω περιορισμοί οπτικοποιούνται στα παρακάτω παραδείγματα:



Σχήμα 3.10: Ευαισθησία στην αρχική θέση των σπόρων, 2 classes, CT image.



Σχήμα 3.11: Ευαισθησία σε outliers, 2 classes, CT image.

Πλεονεκτήματα του παραδοσιακού Random Walker:

1. Η στενή σύνδεση της μεθόδου με τα ηλεκτρικά κυκλώματα, εξασφαλίζει ότι η λύση η οποία παίρνουμε είναι μοναδική, αποτελείται από συνεκτικά τμήματα και είναι μερικώς εύρωστη στο θόρυβο.
2. Το γεγονός ότι η μέθοδος είναι με επίβλεψη συμβάλλει στη διαδραστική εφαρμογή του, όπου ο χρήστης μπορεί να καθορίζει κάποια αρχικά σημεία, να τα μεταβάλλει αν δεν μένει ικανοποιημένος από το αποτέλεσμα και γενικά να ενισχύεται ο ρόλος του ανθρώπινου παράγοντα με σκοπό την επιτυχή εκτέλεση της εφαρμογής.
3. Προκύπτουν αρκετά καλά ποσοστά επιτυχίας από τη μέθοδο αυτή (και τις αντίστοιχες παραλλαγές αυτού), τα οποία μπορούν να βελτιωθούν περισσότερο αν χρησιμοποιήσει κανείς πιο ισχυρά χαρακτηριστικά ή συνδυάσει τον αλγόριθμο με άλλες τεχνικές προεπεξεργασίας και μετα-επεξεργασίας.
4. Η επεκτασιμότητα του βασικού αλγόριθμου επιτρέπει τη χρήση του σε πληθώρα εφαρμογών με μικρές τροποποιήσεις. Ενδεικτικές εφαρμογές αποτελούν η 3D κατάτμηση και η ιατρική απεικόνιση.

3.3.6 Unbiased Random Walker

Η στενή σύνδεση του αλγορίθμου με την επίλυση της εξίσωσης θερμότητας δίνει το έναυσμα να εκμεταλλευτεί κανείς γνωστά αποτελέσματα από την εξίσωση αυτή. Έχουμε ήδη αναφέρει ότι η ελαχιστοποίηση του συναρτησιακού $J_{RW}(x)$, $x \in \mathbf{R}^N$ πραγματοποιείται μέσω του πίνακα L , ο οποίος παρά τη διαδεδομένη μορφή του, υπονοεί ότι η μεταβολή επί μιας ακμής e_{ij} καθορίζεται από τη διαφορά $w_{ij}(x_i - x_j)$. Η μορφή αυτή διατηρεί τη συμμετρία $\partial_j x_i = -\partial_i x_j$, αλλά δε λαμβάνει υπόψη το διαφορετικό βαθμό των κόμβων i, j . Για μια διαδοχή χρονικών στιγμών $[t - 1, t, t + 1]$ έχουμε:

$$\Delta I_{i,t+1} = \sum_{j \sim i} \frac{w_{ij}}{d_i} (x_{i,t} - x_{j,t})$$

$$x_{i,t} = x_{i,t-1} - \Delta I_{i,t}$$

$$x_{j,t} = x_{j,t-1} - \Delta I_{j,t}$$

$$\Delta I_{j,t} = \sum_{k \sim j} \frac{w_{jk}}{d_j} (x_{j,t} - x_{k,t})$$

δηλαδή η μεταβολή της πληροφορίας στον κόμβο i εξαρτάται από τις μεταβολές της πληροφορίας των γειτόνων (και των γειτόνων αυτών κλπ) οι οποίες δεν είναι ίδιες ποιοτικά λόγω του διαφορετικού βαθμού που έχουν οι κόμβοι μεταξύ τους. Επομένως, εκμεταλλευόμαστε το στοιχείο αυτό για να εισάγουμε την ελαχιστοποίηση του

$$J(x) = \frac{1}{2} x^\top L_n x$$

και στον αλγόριθμο του Τυχαίου Περιπατητή. Επειδή ακριβώς κανονικοποιούμε την πληροφορία και την μεταβολή αυτής ανά timestep ονομάζουμε αυτήν την εκδοχή του αλγορίθμου Unbiased Random Walker ή Normalized Random Walker. Σημειώνουμε ακόμα μια πιθανή αναμόρφωση της ανάλυσης. Αν ορίσουμε συνάρτηση $\epsilon = f(d_i, d_j)$ με $0 < \epsilon < |d_i - d_j|$ και αντίστοιχη μόνιμη κατάσταση για έναν κόμβο i :

$$x_i = \frac{1}{\sqrt{d_i}} \sum_{j \sim i} \frac{w_{ij}}{\sqrt{\min(d_i, d_j) + \epsilon}} I_{j,t} = \frac{1}{\sqrt{d_i}} \sum_{j \sim i} \frac{w_{ij}}{\sqrt{\kappa_{i,j}}} I_{j,t} \text{ οπότε έχουμε τις περιπτώσεις:}$$

1. $d_i > d_j \Rightarrow \kappa_{i,j} \in [d_j, d_i]$
2. $d_i = d_j \Rightarrow \kappa_{i,j} = d_i$
3. $d_i < d_j \Rightarrow \kappa_{i,j} \in [d_i, d_j]$

οπότε προκύπτουν εξισώσεις οι οποίες παρεμβάλλονται μεταξύ του παραδοσιακού Random Walker και του Unbiased.

3.3.7 Περιορισμοί και πλεονεκτήματα του Unbiased Random Walker

Περιορισμοί του Unbiased Random Walker:

1. Αναπόδραστα, έχουν κληρονομηθεί τα αρνητικά χαρακτηριστικά του κλασσικού αλγορίθμου, όπως η ευαισθησία στην αρχική θέση των σπόρων και η ανάγκη προσδιορισμού του σ .
2. Για όλους τους κόμβους όπου έστω και ένας γείτονας τους δεν έχει τον ίδιο βαθμό μαζί τους, έπαψε να ισχύει η αρμονική ιδιότητα (harmonic property). Με τον τρόπο αυτό, αναιρείται η αυστηρή έννοια της πιθανότητας όπως αυτή χρησιμοποιείται, δηλαδή αν ανθροίσει κανείς τις ποσότητες οι οποίες προκύπτουν από τη λύση του νέου προβλήματος, τότε δεν προκύπτει μονάδα. Η παραβίαση αυτή ίσως έχει κάποιες επιπτώσεις τις οποίες χρειάζεται να μελετήσει κανείς προτού αξιολογηθεί βέβαιη η υπεροχή της Unbiased εκδοχής.

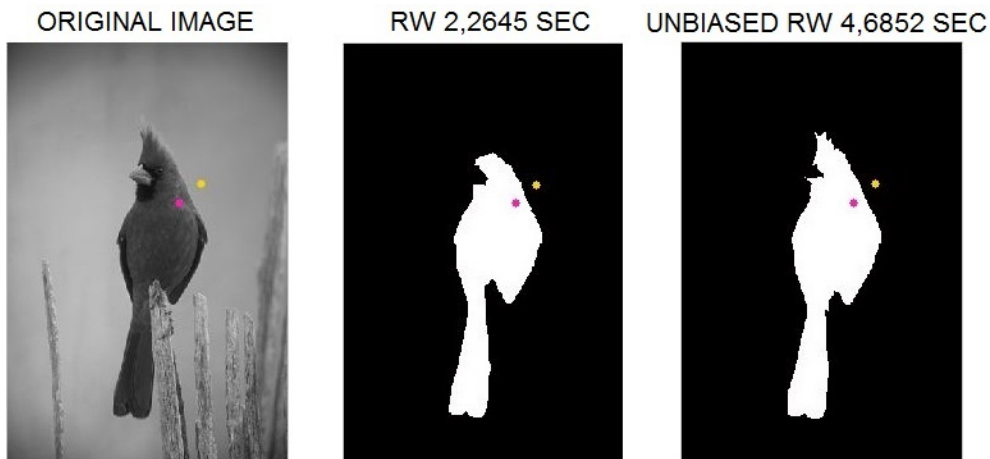
Πλεονεκτήματα του Unbiased Random Walker:

1. Ενσωματώσαμε με φυσικό τρόπο μια παραλλαγμένη μερική διαφορική εξίσωση, η οποία προήλθε από το μοντέλο SIR και η οποία δε διατηρεί πλέον τη συμμετρία $\partial_j x_i = -\partial_i x_j$ αλλά ικανοποιεί μια άλλη ενδιαφέρουσα ιδιότητα της μορφής $\sqrt{d_i} \partial_j x_i = -\sqrt{d_j} \partial_i x_j$, η οποία διατηρεί τη συμμετρία στην ειδικότερη περίπτωση όπου: $\sqrt{d_j} = \sqrt{d_i}$ δηλαδή για $d_j = d_i$.
2. Σε σύγκριση με την κλασσική εκδοχή του αλγορίθμου (για λόγους ισονομίας αναφερόμαστε στην εκδοχή που χρησιμοποιεί το Region Adjacency γράφο), φαίνεται να υπάρχει μια βελτίωση σε πολλές περιπτώσεις, είτε όταν υπάρχει κάποιο outlier είτε όχι. Οι περιπτώσεις όπου υστερούσε ο Unbiased Random Walker ήταν πολύ λιγότερες και κυρίως ήταν περιπτώσεις όπου υπήρχαν μικροδιαφορές.

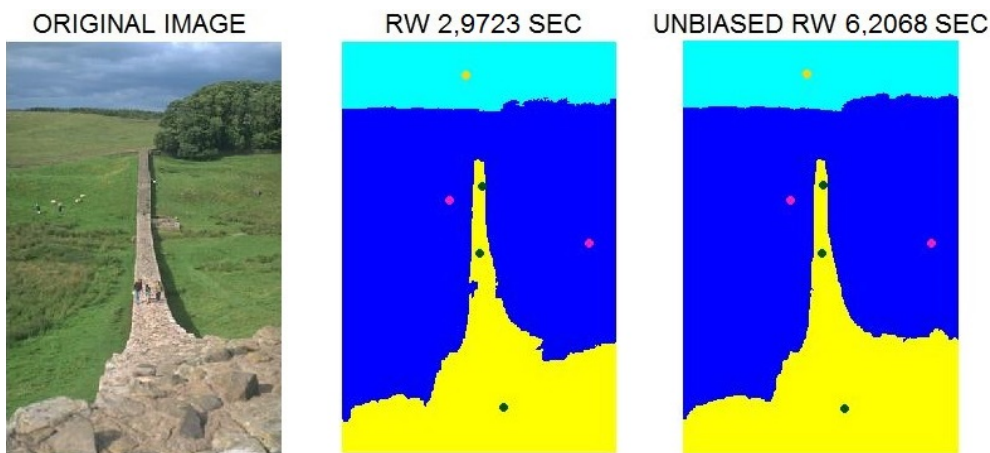
3. Παρατηρούμε ότι η ανάλυση που κάναμε (δίχως δηλαδή το 4/8 regular grid), δίνει άμεση επιτάχυνση του παραδοσιακού αλγορίθμου επιτρέποντας την ακόμη πιο γρήγορη κατάτμηση μικρών εικόνων, αλλά και την κατάτμηση μεγαλύτερων από 256×256 σε χρόνο πολύ μικρότερο από ότι θα χρειαζόταν σε διαφορετική περίπτωση. Μάλιστα, όπως θα φανεί και στην επόμενη ενότητα, η χρήση RAG δίνει πολύ καλύτερα αποτελέσματα για τον Unbiased Random Walker και μειωμένη ευαισθησία στην παράμετρο σ_{kernel} αλλά και στους αρχικούς σπόρους - seeds.

3.3.8 Οπτικοποίηση-Αποτελέσματα

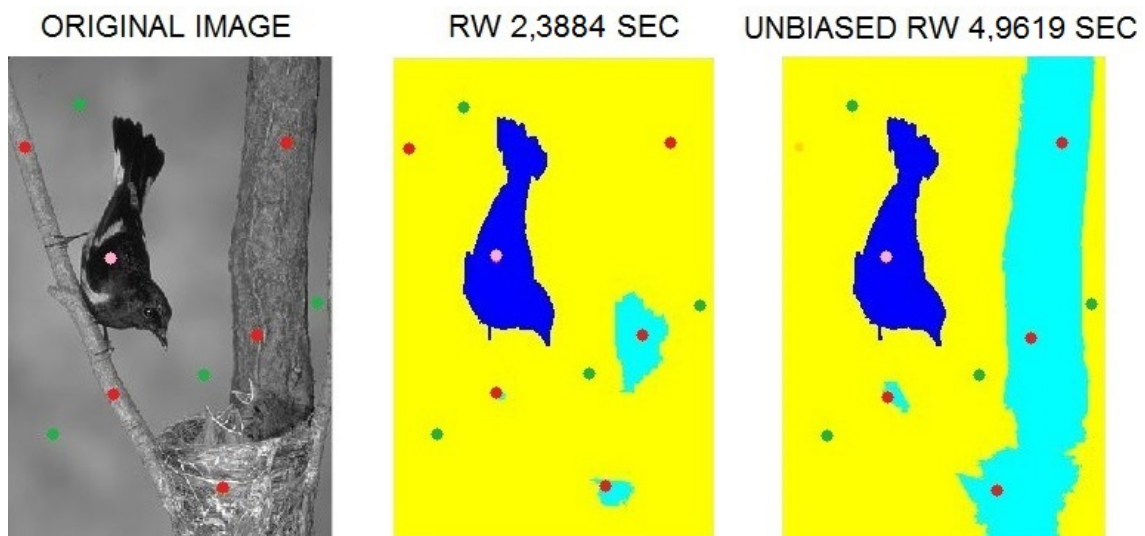
Στην ενότητα αυτή, θα οπτικοποιηθούν τα αποτελέσματα των 2 αλγορίθμων (παραδοσιακός Random Walker και Unbiased Random Walker) και για την περίπτωση όπου χρησιμοποιεί κανείς 4/8 regular grid) αλλά και για Region Adjacency γράφο. Επιπλέον, θα εξεταστούν οι αποδόσεις των αλγορίθμων καθώς μεταβάλλει κανείς την παράμετρο σ_{kernel} αλλά και τους αρχικούς σπόρους-seeds.



Σχήμα 3.12: Regular 8 Grid, $\sigma_{kernel} = 90$, 2 classes, bird gray.



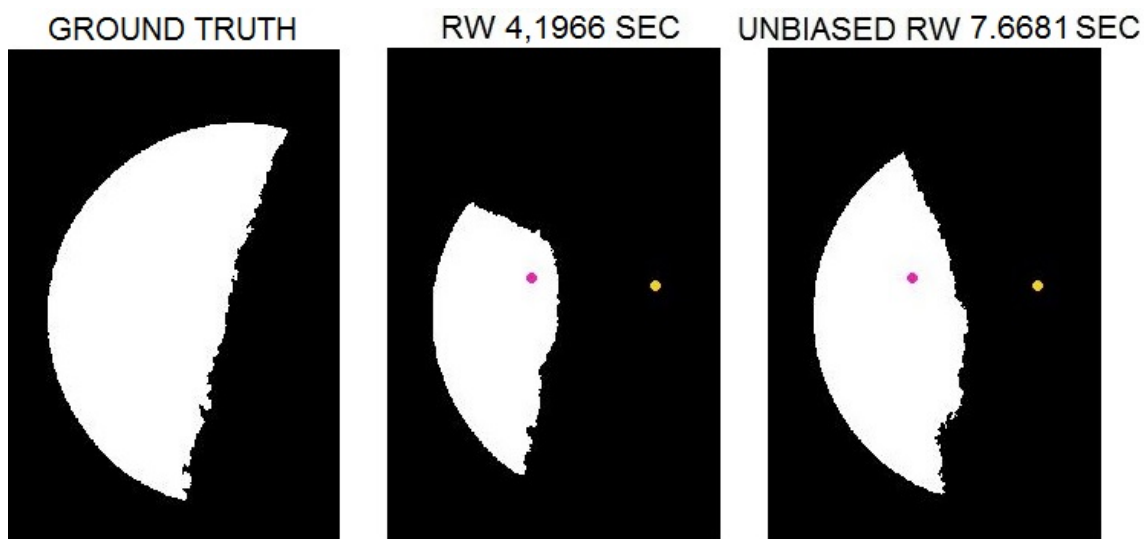
Σχήμα 3.13: Regular 8 Grid, $\sigma_{kernel} = 90$, 3 classes, road.



Σχήμα 3.14: Regular 8 Grid, $\sigma_{kernel} = 90$, 3 classes, small bird gray.



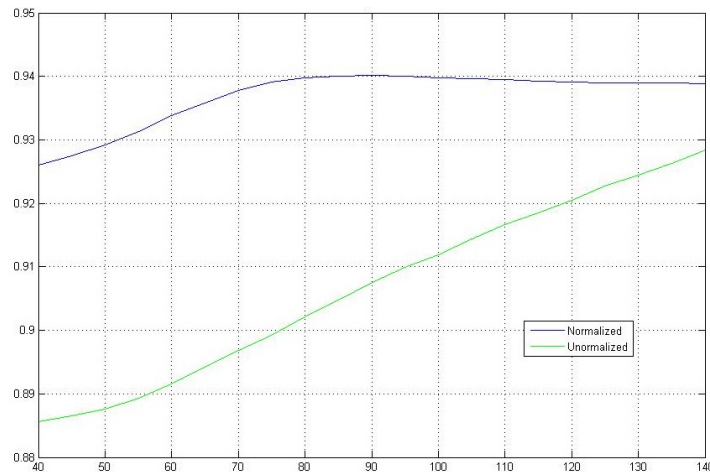
Σχήμα 3.15: Regular 8 Grid, $\sigma_{kernel} = 90$, 2 classes, swan gray.



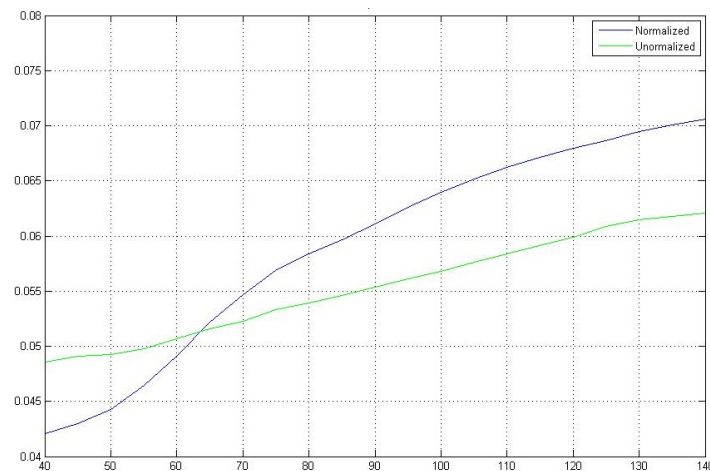
Σχήμα 3.16: Ground Truth, Regular 8 Grid, $\sigma_{kernel} = 90$, 2 classes, moon.

Με βάση την παραπάνω εικόνα και το αντίστοιχο ground truth segmentation, μεταβάλλουμε την παράμετρο σ_{kernel} στο διάστημα $[40, 140]$ με βήμα 5 και για κάθε μια τιμή του σ_{kernel} μετράμε το F-score σε κάθε μία από τις 5 αρχικοποιήσεις σπόρων που έγιναν χειροκίνητα. Θυμίζεται ότι το F-score μας δείχνει την ποιότητα της τελικής κατάτμησης με βάση το ground truth segmentation. Έπειτα, υπολογίζουμε τη μέση τιμή για κάθε τιμή του σ_{kernel} καθώς και το standard deviation.

Από τα παρακάτω διαγράμματα, παρατηρούμε ότι όταν $\sigma_{kernel} \in [40, 140]$ τότε ο Unbiased Random Walker έχει μια αυξημένη απόδοση κατά περίπου 3% μέχρι περίπου την τιμή $\sigma_{kernel} = 100$. Για τιμές μεγαλύτερες, η βελτίωση του Unbiased Random Walker μειώνεται στο 1%, δίνοντας πάντα καλύτερη μέση τιμή στο τελικό αποτέλεσμα. Ταυτόχρονα, παρατηρούμε ότι οι τυπικές αποκλίσεις είναι πολύ κοντά μεταξύ τους (η μεγαλύτερη διαφορά παρατηρείται στην τιμή 0.01) και μικρά σε τιμή το καθένα, οπότε τα προηγούμενα αποτελέσματα έχουν σχετικά μικρές αποκλίσεις. Συμπερασματικά, εξάγουμε το συμπέρασμα ότι ο Unbiased Random Walker επηρεάζεται λιγότερο από ότι ο παραδοσιακός αλγόριθμος και μπορεί να δώσει καλύτερα αποτελέσματα. Η υπεροχή του αλγορίθμου θα φανεί κύρια με τη χρήση RAG.

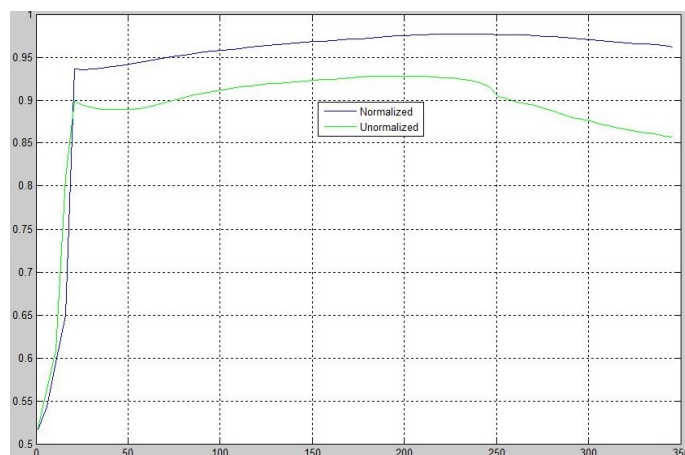


Σχήμα 3.17: Regular 8 Grid, 5 seed types, $\sigma_{kernel} \in [40, 140]$, F-score mean.



Σχήμα 3.18: Regular 8 Grid, 5 seed types, $\sigma_{kernel} \in [40, 140]$, F-score standard deviation.

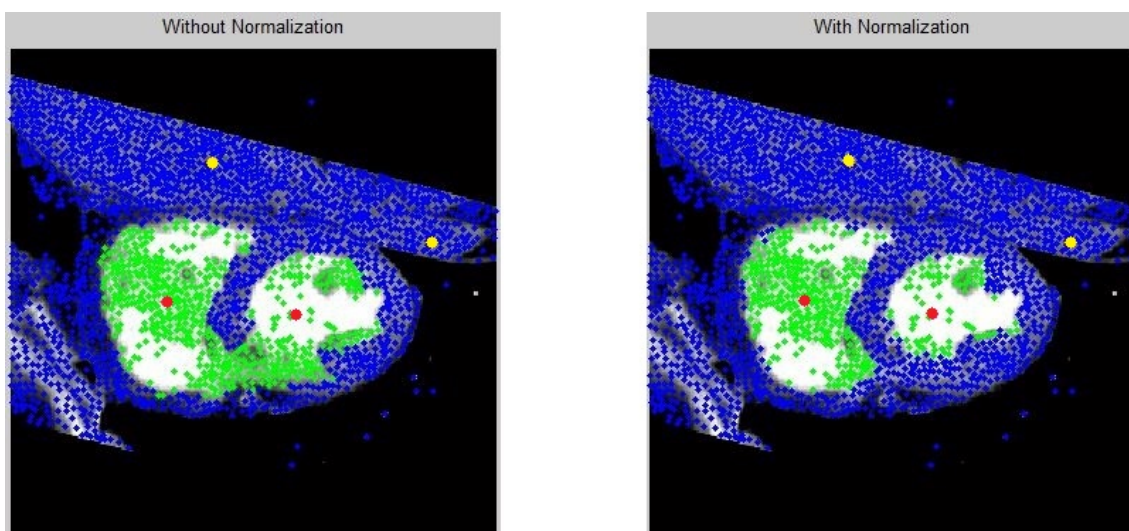
Στο προηγούμενο συμπέρασμα μας οδήγησε και το επόμενο πείραμα: μεταβάλλοντας την παράμετρο σ_{kernel} σε ένα πολύ πιο ευρύ διάστημα ($[1,350]$ με βήμα 5) και για 2 διαφορετικές θέσεις αρχικών σπόρων, παρατηρούμε εκ νέου την υπεροχή του Unbiased Random Walker.



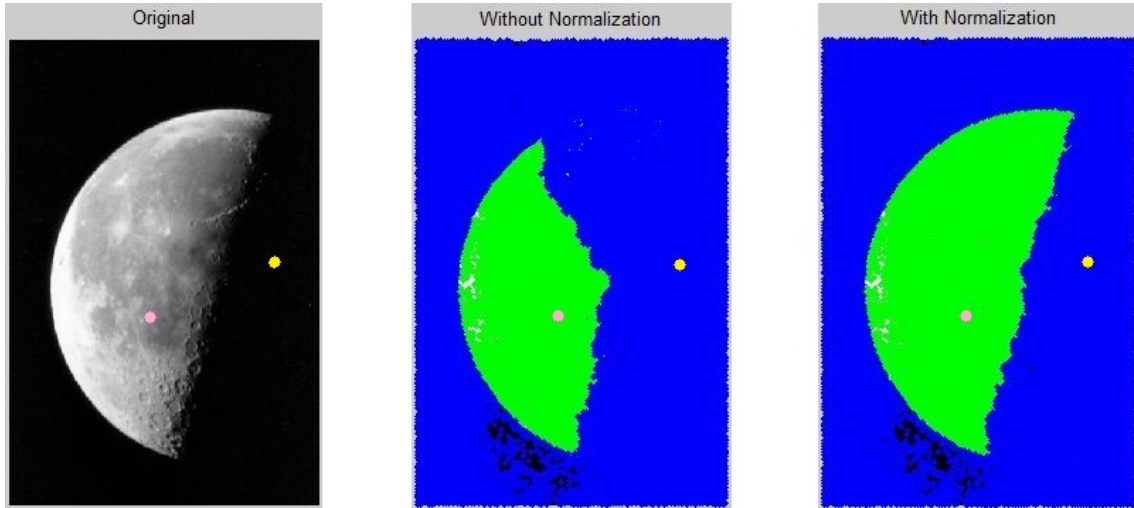
Σχήμα 3.19: Regular 8 Grid, 10 seed types, $\sigma_{kernel} \in [1, 350]$, F-score mean.

Στη συνέχεια, επικεντρωνόμαστε στην περίπτωση του Region Adjacency γράφου, όπου παρατηρούμε ξανά την υπεροχή του Unbiased Random Walker, κάποιες φορές και εντονότερη από ότι πριν. Ο λόγος πίσω από αυτήν την παρατήρηση είναι το γεγονός ότι πλέον οι γείτονες κάθε κόμβου δεν είναι σταθεροί σε αριθμό (4 ή 8) αλλά ποικίλλουν ανάλογα με την εικόνα και την εκάστοτε γειτονιά των κόμβων.

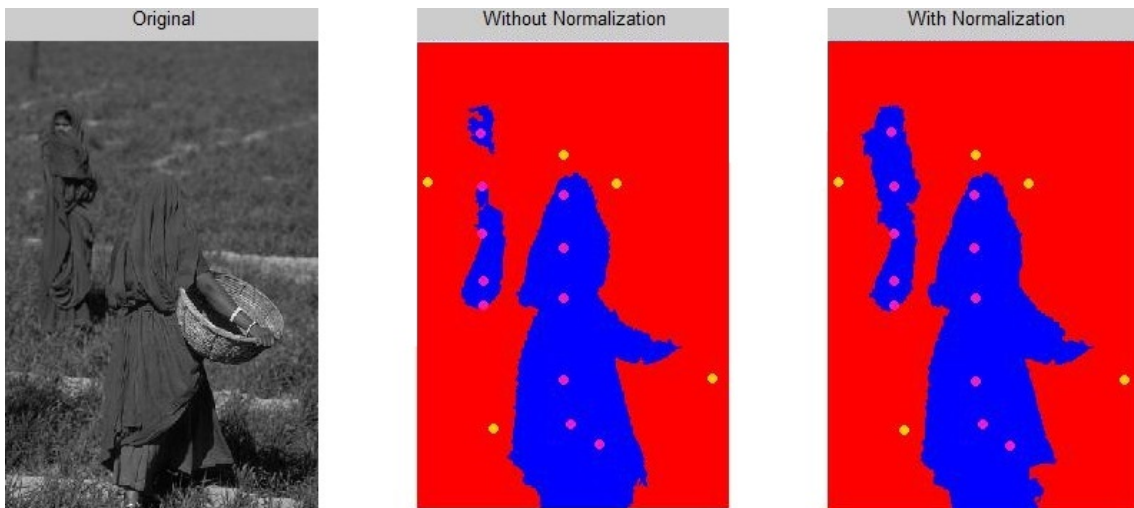
Από τα παρακάτω διαγράμματα, παρατηρούμε ότι ακόμα και σε περιπτώσεις όπου στο regular 8 grid ο Unbiased Random Walker είχε χειρότερη επίδοση (όπως στο παράδειγμα του κύκνου), τώρα διατηρεί το αποτέλεσμα της κατάτμησης πολύ καλύτερα (σε σχέση με τον παραδοσιακό αλγόριθμο) και όταν μεταβάλλουμε την παράμετρο σ_{kernel} .



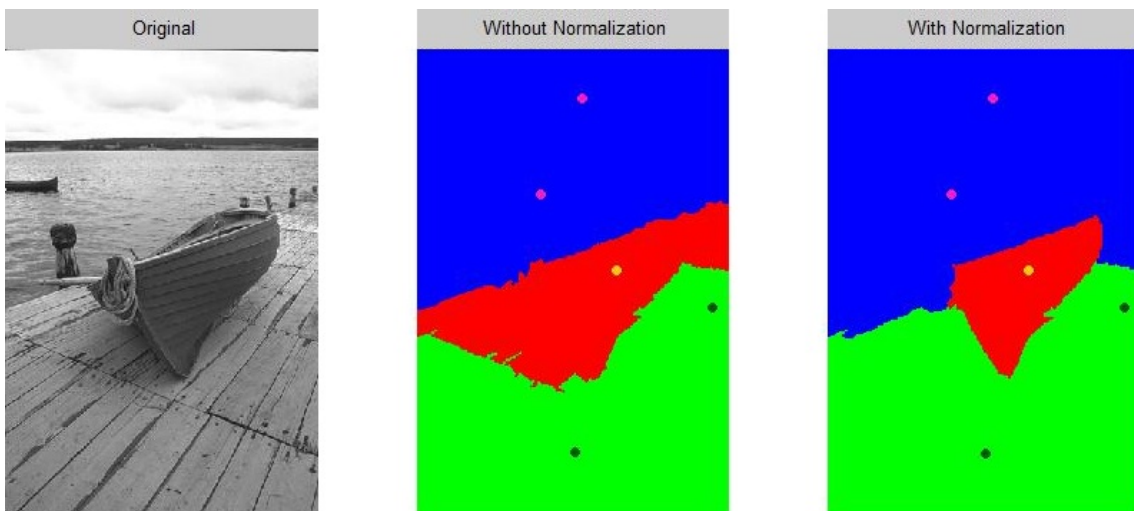
Σχήμα 3.20: Region Adjacency γράφος $\sigma_{kernel} = 90$, 2 classes, CT image.



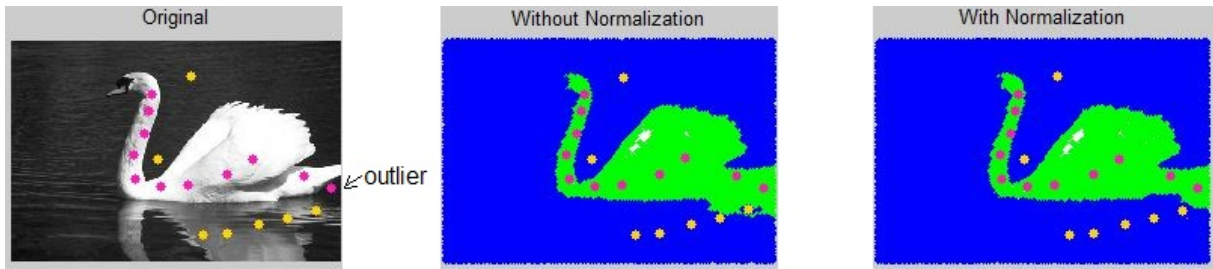
Σχήμα 3.21: Region Adjacency γράφος $\sigma_{kernel} = 90$, 2 classes, moon.



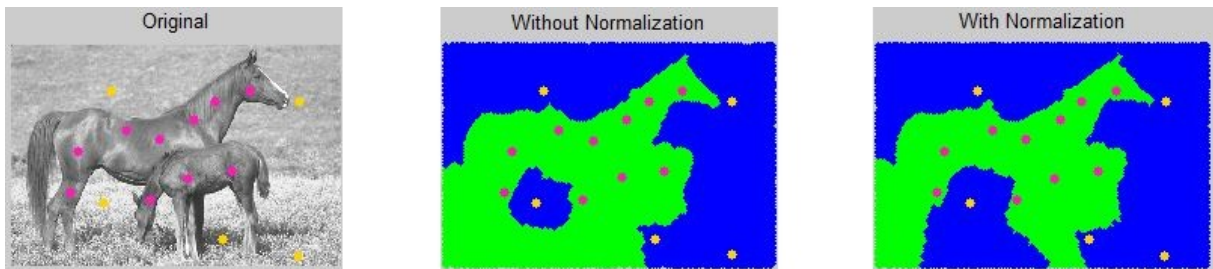
Σχήμα 3.22: Region Adjacency γράφος $\sigma_{kernel} = 90$, 2 classes, women gray.



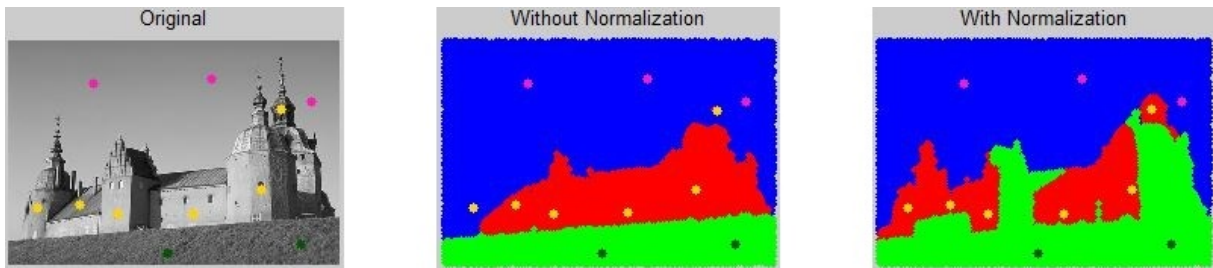
Σχήμα 3.23: Region Adjacency γράφος $\sigma_{kernel} = 90$, 2 classes, boat gray.



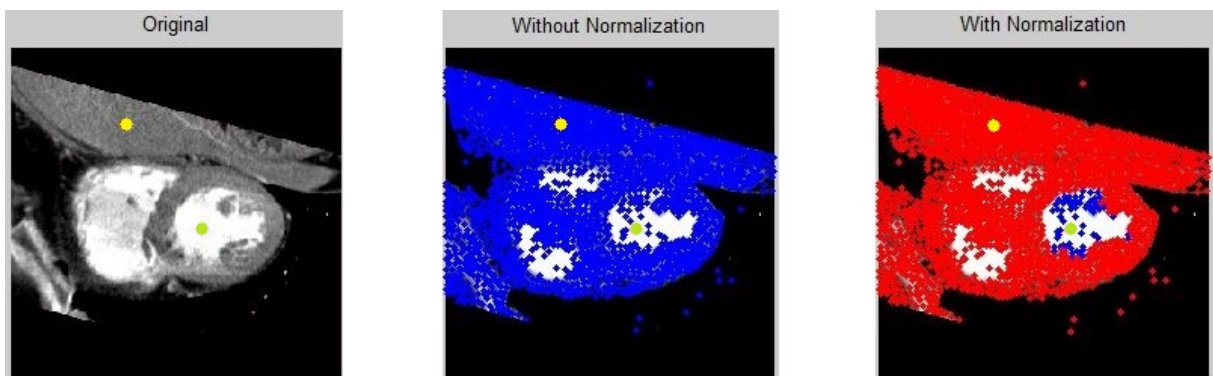
Σχήμα 3.24: Region Adjacency γράφος $\sigma_{kernel} = 90$ + outliers, 2 classes, swan gray.



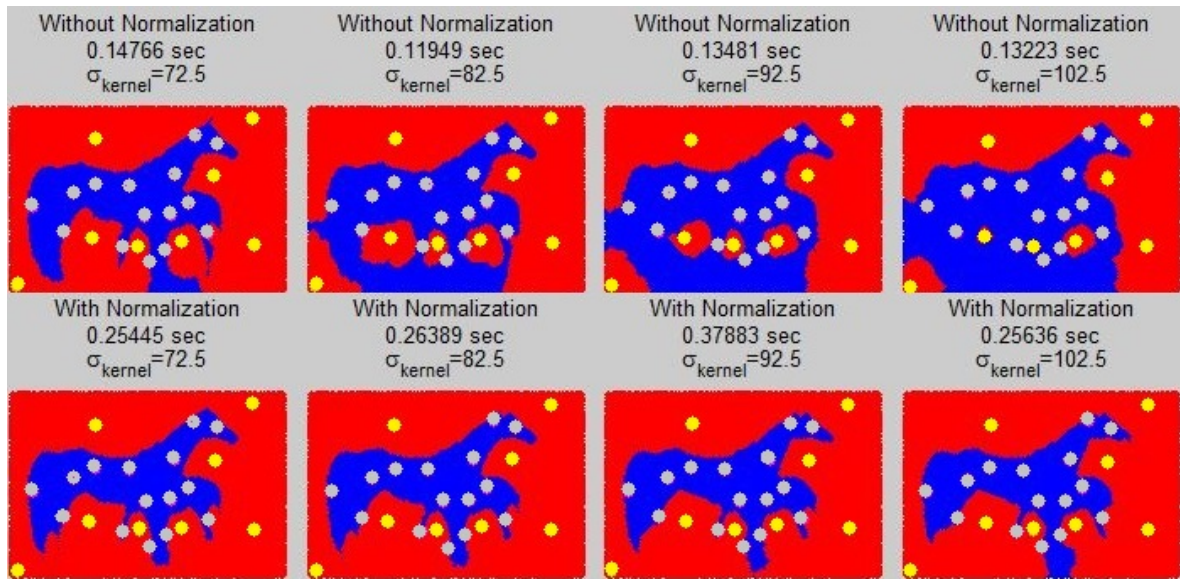
Σχήμα 3.25: Region Adjacency γράφος $\sigma_{kernel} = 90$, 2 classes, horses gray.



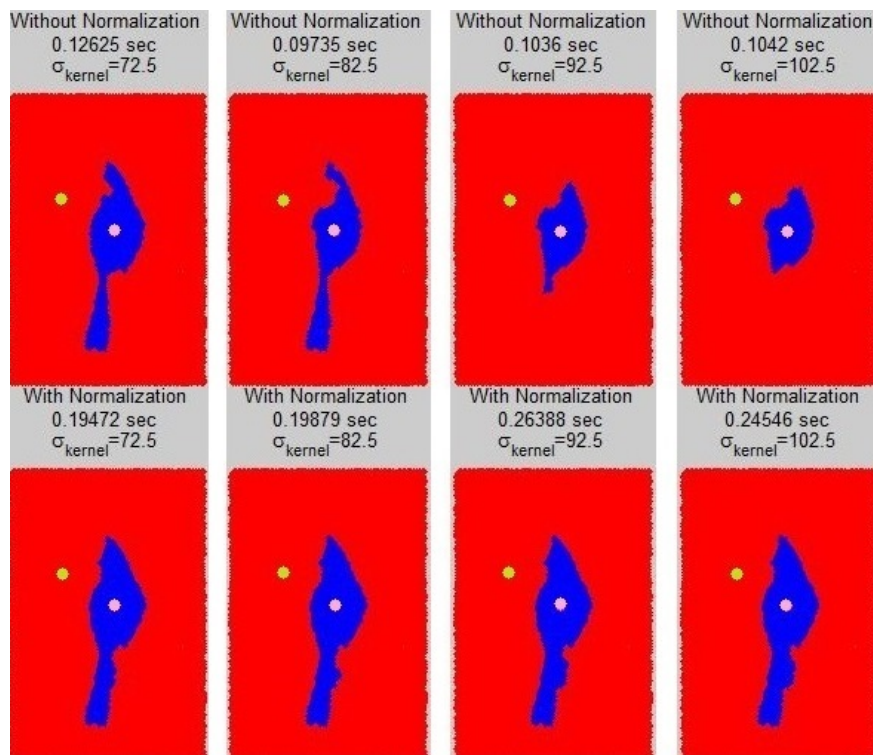
Σχήμα 3.26: Region Adjacency γράφος $\sigma_{kernel} = 90$, 3 classes, castle gray.



Σχήμα 3.27: Region Adjacency γράφος, Unbiased Outperforms, seeds= [1672, 1024], 2 classes, CT image.



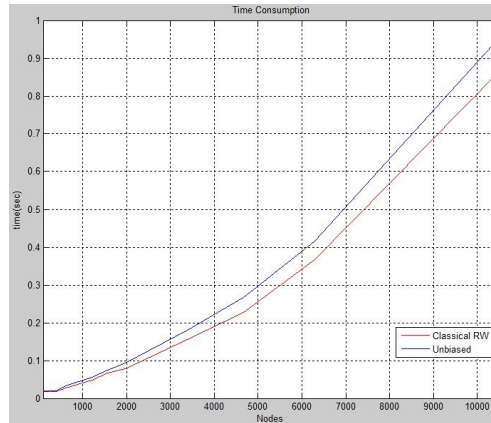
Σχήμα 3.28: Region Adjacency γράφος, Unbiased Outperforms, 2 classes, horses (using Lab).



Σχήμα 3.29: Region Adjacency γράφος, Unbiased Outperforms, 2 classes, bird gray.

3.3.9 Υπολογιστικό κόστος για Unbiased Random Walker

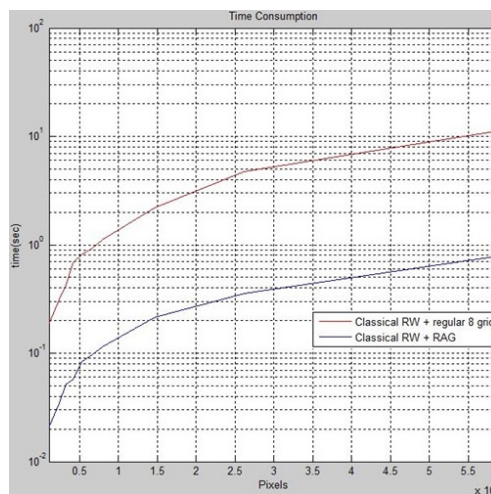
Η πρόσθετη υπολογιστική επιβάρυνση η οποία αντιστοιχεί στον Unbiased Random Walker θα οφείλεται αποκλειστικά στον υπολογισμό του $D^{-\frac{1}{2}}$ και τον πολλαπλασιασμό του $D-W$ από δεξιά και από αριστερά μαζί του. Όπως παρατηρούμε και στο κάτωθι διάγραμμα, η επιβάρυνση αυτή είναι μικρή και δεν αυξάνεται σημαντικά με την αύξηση των κόμβων.



Σχήμα 3.30: Υπολογιστικοί χρόνοι ανά χιλιάδα κόμβων.

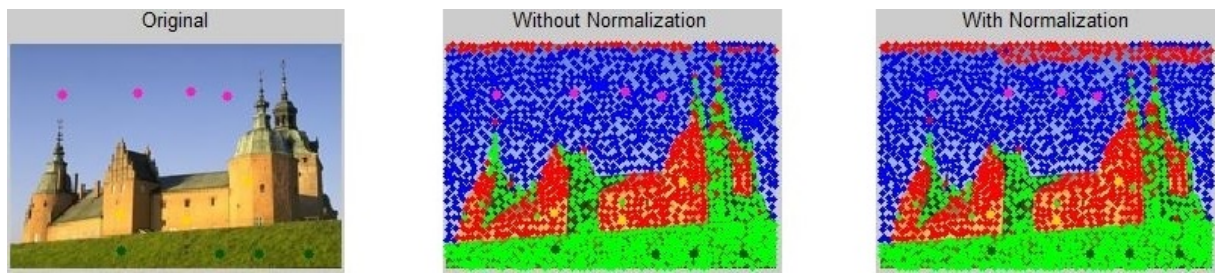
3.3.10 Υπολογιστικό κέρδος RAG vs Regular Grid

Ως τώρα, δεν εστιάσαμε στη σημασία της χρήσης ενός Region Adjacency γράφου έναντι ενός κανονικού πλέγματος. Με βάση το επόμενο διάγραμμα, βλέπουμε μια πολύ μεγάλη διαφορά στους χρόνους εκτέλεσης συναρτήσεων των pixel. Συγκεκριμένα, παρατηρούμε ότι για περίπου 6×10^5 pixel απαιτούνται περίπου 12 sec για ένα κανονικό πλέγμα 8 γειτόνων έναντι περίπου 0.8 sec στην περίπτωση ενός Region Adjacency γράφου. Σημειώνουμε, επομένως, ότι δίχως σημαντική μεταβολή στην τελική κατάτμηση της εικόνας, μειώνουμε δραματικά τον αριθμό των κόμβων και των ακμών του γράφου \mathcal{G} και συνάμα τους απαιτούμενους υπολογιστικούς πόρους.

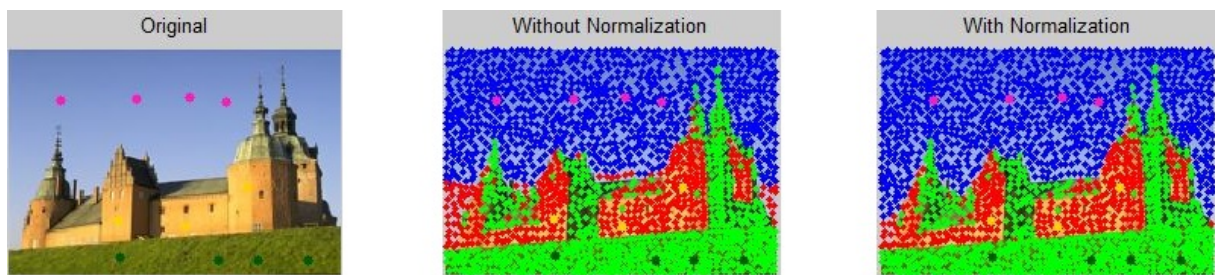


Σχήμα 3.31: Υπολογιστικοί χρόνοι ανά χιλιάδα pixel.

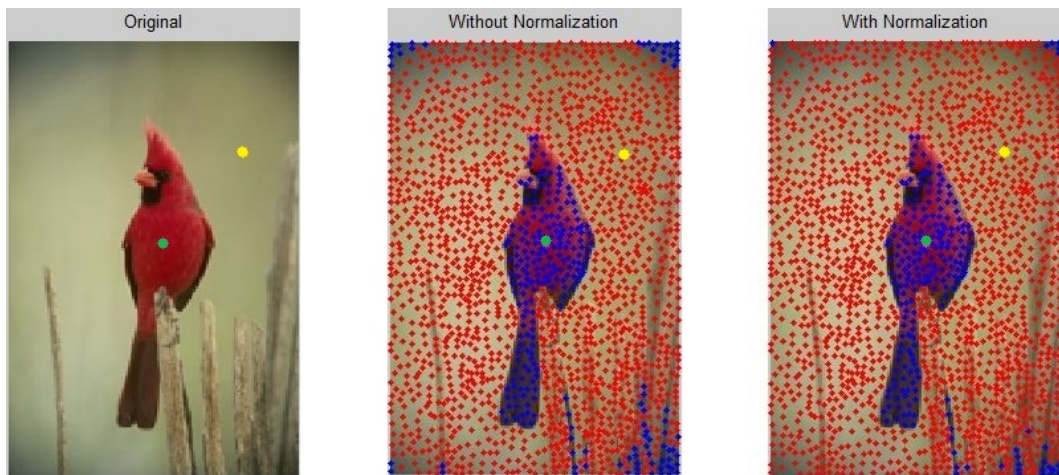
3.3.11 Συνδυάζοντας το LNP με τον αλγόριθμο Τυχαίου Περιπατητή



Σχήμα 3.32: Un-normalized vs Normalized + LNP, $k = 5$, 3 classes, castle.



Σχήμα 3.33: Un-normalized vs Normalized + LNP, $k = 15$, 3 classes, castle.



Σχήμα 3.34: Un-normalized vs Normalized + LNP, $k = 15$, 2 classes, bird.

Παρατηρώντας τα παραπάνω διαγράμματα, βλέπουμε ότι όντως η μέθοδος LNP μπορεί να αποδώσει τις σχέσεις μεταξύ των κόμβων με βάση την ελαχιστοποίηση που προτείνεται, δίχως την ανάγκη προσδιορισμού της ευαισθητής ως προς το αποτέλεσμα παραμέτρου σ του Gaussian πυρήνα. Ταυτόχρονα, όμως χρειάζεται πάλι να επιλέξουμε την παράμετρο k οι οποία αντιστοιχεί στον αριθμό των γειτόνων για κάθε κόμβο, παράμετρος η οποία όπως βλέπουμε, δε μπορεί να τίθεται αυτόματα σε κάποια σταθερή τιμή για όλες τις εικόνες. Επίσης, η επίλυση του προβλήματος ελαχιστοποίησης είναι υπολογιστικά επιβαρυντική ακόμα και για μικρό αριθμό κόμβων (~ 2000) και καθώς αυξάνει κανείς την παράμετρο k . Σημειώνουμε επίσης, ότι ο πίνακας X που χρησιμοποιήθηκε στην ελαχιστοποίηση, περιείχε μόνο τις μέσες φωτεινότητες των κόμβων και όχι τις χωρικές συντεταγμένες αυτών.

3.3.12 Regularization Framework

Ως τώρα αντιμετωπίσαμε μονάχα εφαρμογές της διάχυσης θερμότητας στο γράφο με σκοπό την κατάτμηση αντικειμένων. Μια άλλη χρήση παρόμοιων ιδεών είναι και το regularization. Στο [30] γίνεται έντονη προσπάθεια για την ενσωμάτωση ενός ενιαίου regularization framework με σκοπό το smoothing και το denoising μιας εικόνας η οποία αντιστοιχίζεται με ένα γράφο $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Έστω ότι ο γράφος είναι συνδεδεμένος, με βάρη τα οποία είναι συμμετρικά και μη κατευθυνόμενος. Ακόμα, ορίζουμε το χώρο $\mathcal{H}(\mathcal{V})$ ως το χώρο Hilbert των πραγματικών συναρτήσεων $f : \mathcal{V} \rightarrow \mathcal{R}^m$ οι οποίες αναθέτουν ένα διάνυσμα $f(v)$ σε κάθε κόμβο. Ο χώρος αυτός είναι εφοδιασμένος με την πράξη του εσωτερικού γινομένου:

$$\langle f, h \rangle_{\mathcal{H}(\mathcal{V})} = \sum_{v \in \mathcal{V}} f(v)h(v)$$

Ομοίως, μπορεί κανείς να ορίσει ένα χώρο $\mathcal{H}(\mathcal{E})$ των πραγματικών συναρτήσεων επί των ακμών \mathcal{E} . Επίσης, ορίζονται τα εξής:

$$\text{κατευθυνόμενη παράγωγος: } \partial_v f(u) = \sqrt{w(u, v)(f(v) - f(u))} \quad \forall u, v \in \mathcal{E}$$

$$\text{τελεστής weighted gradient: } \nabla_w f(v) = (\partial_v f(u) : u \sim v)^\top$$

$$\text{local variation: } \|\nabla_w f(v)\| = \sqrt{\sum_{u \sim v} (\partial_v f(u))^2} = \sqrt{\sum_{u \sim v} w(u, v)(f(v) - f(u))^2}$$

$$\text{τελεστής weighted Laplace: } \Delta_w^2 f(v) = \sum_{u \sim v} w(u, v)(f(v) - f(u))$$

Έπειτα, το πρόβλημα του regularization τίθεται ως εξής: δοθείσης μια αρχικής συνάρτησης f_0 επί των κόμβων ενός γράφου \mathcal{G} , να προσδιοριστεί μια συνάρτηση f^* η οποία να είναι smooth με βάση το $\Delta_w^2 f(v)$ αλλά και να είναι κοντά στην αρχική συνάρτηση. Με άλλα λόγια:

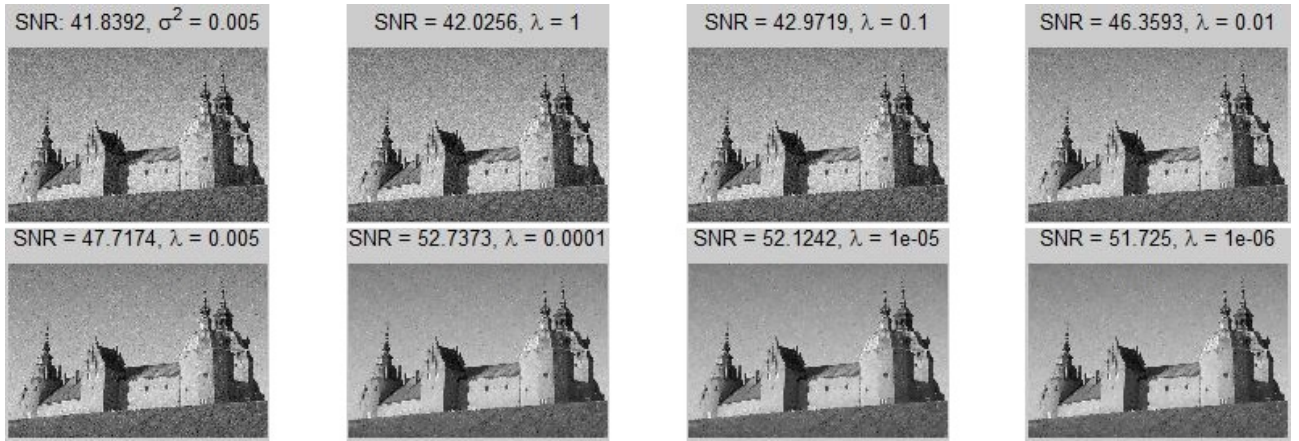
$$f^* = \min_{f \in \mathcal{V}} E_w^2(f, f^0, \lambda) := \frac{1}{2} \sum_{v \in \mathcal{V}} \|\nabla_w f(v)\|^2 + \frac{\lambda}{2} \|f - f^0\|_{\mathcal{H}(\mathcal{V})}^2$$

όπου λ είναι μια παράμετρος που εκφράζει το tradeoff μεταξύ των 2 όρων. Η λύση του προβλήματος είναι μοναδική και προκύπτει επαναληπτικά ως:

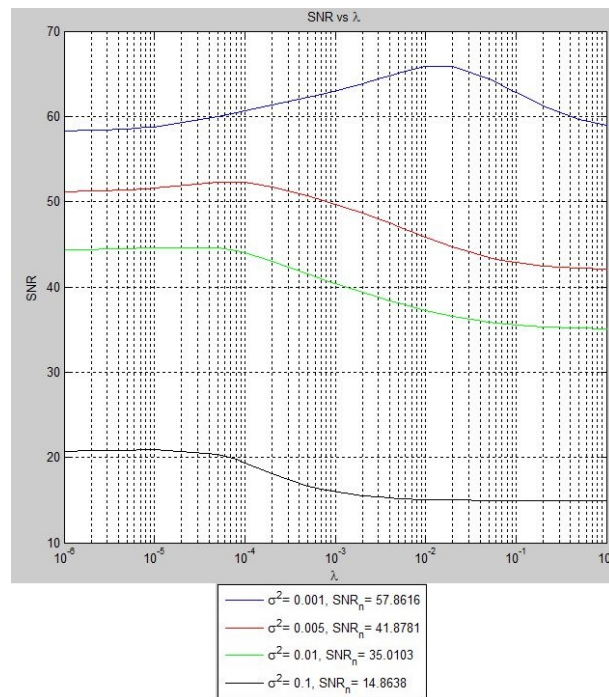
$$f^{t+1}(v) = \phi^t(v, v)f^0(v) + \sum_{u \sim v} \phi^t(v, u)f^t(u) \quad \forall v \in \mathcal{V}$$

$$\text{με } \phi^t(v, u) = \phi(v, u) = \frac{2w(u, v)}{2\lambda + \sum_{u \sim v} 2w(u, v)}, \phi^t(v, v) = \phi(v, v) = \frac{2\lambda}{2\lambda + \sum_{u \sim v} 2w(u, v)}$$

Οι εξισώσεις αυτές μοιάζουν με τις αντίστοιχες του diffusion αφού χρησιμοποιήθηκαν παρόμοιες ιδέες. Όμως, σε εκείνη την περίπτωση είχαμε τη διάδοση μιας ασθένειας επί του γράφου, ενώ τώρα ομαλοποιούμε τις τιμές τις οποίες λαμβάνει μια συνάρτηση επί των κόμβων του γράφου ώστε να πάρουμε μια ομαλοποιημένη εκδοχή της αρχικής εικόνας.



Σχήμα 3.35: Grayscale regularization, $\sigma_{kernel} = 90$, castle gray.



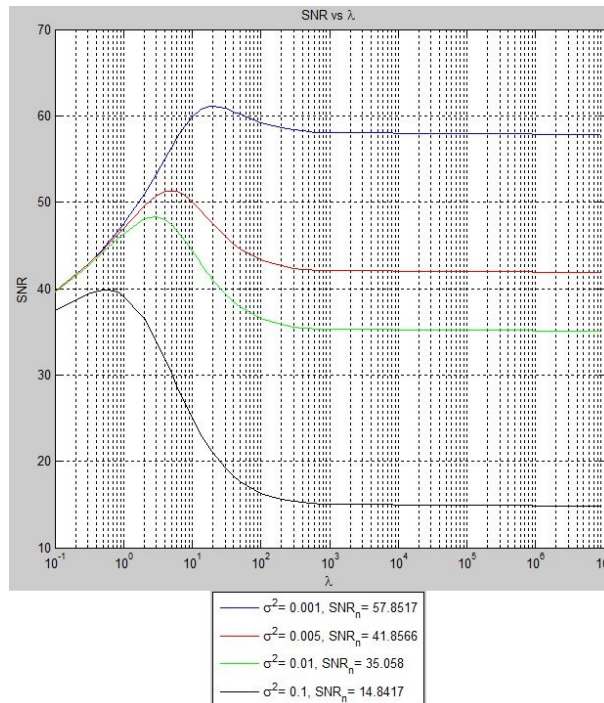
Σχήμα 3.36: SNR Analysis, $\sigma_{kernel} = 90$, castle gray.

Με βάση τα παραπάνω διαγράμματα παρατηρούμε ότι η βελτίωση της ποιότητας της εικόνας φτάνει σε κορεσμό για μεγάλες τιμές της παραμέτρου λ . Επίσης, όσο αυξάνει η μη προβλεψιμότητα του θορύβου, τόσο πιο δύσκολη γίνεται η βελτίωση της εικόνας. Υπενθυμίζεται ότι βελτίωση της ποιότητας της εικόνας αντιστοιχεί σε μεγαλύτερο SNR (Signal to Noise Ratio). Ακόμα, σημειώνεται ότι το regularization το οποίο υλοποιήθηκε, παίρνει σα δεδομένο τον πίνακα $\mathcal{W} = [w_{i,j}]$ καθώς και το σ_{kernel} το οποίο χρειάζεται κανείς για τα βάρη του πίνακα. Αν αυτή η γνώση δεν είναι διαθέσιμη (όπως συνήθως συμβαίνει), μπορούμε να χρησιμοποιήσουμε αντί του πίνακα \mathcal{W} μια παραλλαγμένη μορφή αυτού:

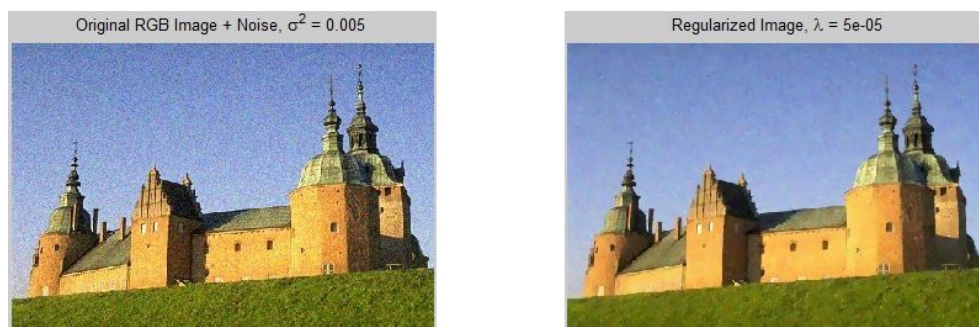
$$w_{i,j} = \begin{cases} 1 & \text{αν } i \sim j \\ 0 & \text{αλλιώς} \end{cases}$$

Από το παρακάτω διάγραμμα παρατηρούμε ότι η βελτίωση της ποιότητας εμφανίζεται για τελείως διαφορετικές τιμές της παραμέτρου λ και επίσης χάνει το βελτιωτικό της χα-

ρακτήρα αρκετά γρήγορα. Όμως, αν κανείς χρησιμοποιήσει τιμές της παραμέτρου λ κοντά στο 20, μπορεί να βελτιώσει την ποιότητα, δίχως τη γνώση του σ_{kernel} . Τέλος, μπορούμε να εφαρμόσουμε το regularization framework και σε έγχρωμες εικόνες, κατ' αναλογία με τα προηγούμενα.



Σχήμα 3.37: SNR Analysis, without σ_{kernel} , castle.



Σχήμα 3.38: RGB regularization, $\sigma_{kernel} = 90$, castle.

Η ανάλυση του regularization framework ακολουθεί κάποιες παρόμοιες εξισώσεις με εκείνες που χρησιμοποιήθηκαν για το diffusion. Θα χρησιμοποιήσουμε τη γνώση αυτή στο επόμενο κεφάλαιο ώστε να εισάγουμε χωρική πληροφορία μέσω αυτού του πίνακα γειτνίασης που ορίσαμε με τη βοήθεια του RAG. Κατ' αναλογία με το regularization όπου κάθε γειτονιά παροτρύνεται να αποκτά παρόμοιες τιμές φωτεινότητας, έτσι θα παραποτρύνονται και γειτονικοί κόμβοι να αποκτούν ίδιες ετικέτες κατά τη διαδικασία του clustering.

3.4 Αλγόριθμος Power Watershed

3.4.1 Περιγραφή του ενιαίου framework

Σε μια προσπάθεια ενσωμάτωσης πολλών γραφοθεωρητικών μεθόδων σε ένα ενιαίο framework προτάθηκε στο [14] ο αλγόριθμος Power Watershed. Αξίζει να αναφερθεί ότι το ενιαίο αυτό πλαίσιο εργασίας περιλαμβάνει πολλές άλλες διαδεδομένες τεχνικές, όπως τα Graph Cuts - GC ([34]) και το μετασχηματισμό Watershed με Minimum Spanning Forest - MSF ([35]). Βασικό χαρακτηριστικό του αλγορίθμου προέρχεται από την ελαχιστοποίηση του ακόλουθου συναρτησιακού:

$$\arg \min_x \sum_{e_{ij} \in E} w_{ij} |x_i - x_j|^q + \sum_{v_i \in V} w_{ij} |x_i - y_i|^q \quad (3.17)$$

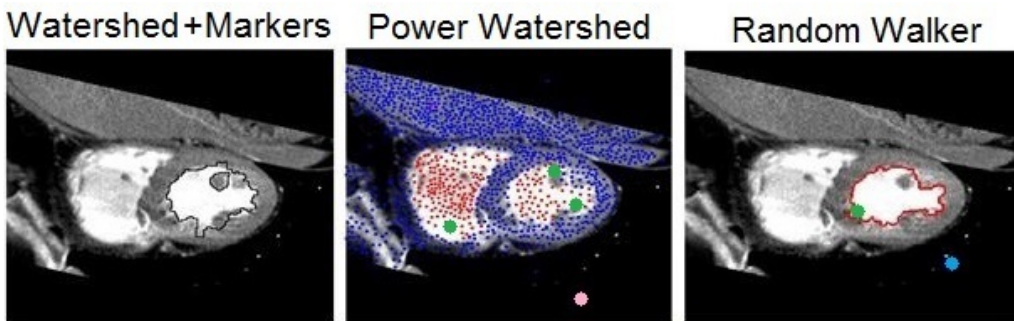
Ο πρώτος όρος αναφέρεται ως όρος smoothness, αποτρέπει δηλαδή απότομες μεταβολές μεταξύ κόμβων που είναι έντονα συσχετισμένοι, ενώ ο δεύτερος είναι ο όρος fidelity δηλαδή η απαίτηση να είμαστε κοντά στις τιμές y_i . Σημειώνεται ότι οι τιμές x_i θα αποτελέσουν μετά απο κατωφλίωση το κριτήριο για την τελική κατάτμηση, ενώ οι y_i είναι δεδομένες. Για απλότητα, θεωρούμε μόνο τον πρώτο όρο και εισάγουμε μια ακόμη παράμετρο p ώστε:

$$\arg \min_x \sum_{e_{ij} \in E} w_{ij}^p |x_i - x_j|^q \quad (3.18)$$

Για διάφορες τιμές των p, q υπάρχουν διάφοροι αλγόριθμοι στους οποίους καταλήγουμε. Για $p \rightarrow +\infty$ και $q = 1, 2$ από την (3.18) παίρνουμε τον αλγόριθμο Power Watershed. Η ελαχιστοποίηση της (1) γίνεται με βάση κάποια constraints ή αλλιώς σημεία σπόρους τα οποία σηματοδοτεί ο χρήστης ως χαρακτηριστικά κάθε περιοχής. Για κατάτμηση ενός μόνο αντικειμένου απαιτούνται 2 ετικέτες (1 για το αντικείμενο και 0 για το background) και οι κόμβοι του γράφου αντιστοιχούν στα pixel της εικόνας (4/8 regular grid). Σημειώνεται ότι για p πεπερασμένο και $q = 2$ η (3.18) μας οδηγεί στον παραδοσιακό Random Walker.

3.4.2 Υλοποίηση σε μη κανονικό πλέγμα

Αντί για το καθιερωμένο regular grid επιλέγονται κόμβοι τα γεωμετρικά κέντρα των περιοχών που προκύπτουν από watershed μετασχηματισμό, περιοχές που αποτελούν υπερκατάτμηση της εικόνας. Η υλοποίηση για $p = 2$ σε μια ιατρική εικόνα έδωσε:



Σχήμα 3.39: Watershed with Markers, Power Watershed, Random Walker, CT image

3.5 Συμπεράσματα κεφαλαίου 3 - Ανακεφαλαίωση

Η πρώτη κύρια συνδρομή του κεφαλαίου αυτού ήταν η ενοποίηση του επιδημιολογικού μοντέλου SIR με ιδέες από label propagation και heat diffusion πάνω σε γράφους. Η βασική επιδίωξη ήταν η ανάλυση των μηχανισμών εξάπλωσης μιας ασθένειας επί των κόμβων του γράφου και η επιτυχής μοντελοποίηση αυτών. Είδαμε ότι στο ανάλογο της μεταφοράς θερμότητας από θερμές περιοχές σε κρύες μπορεί κανείς να αναπαραστήσει τη θερμότητα με ετικέτα η οποία διαχέεται πάνω στο γράφο ή με την εξάπλωση μιας ασθένειας μεταξύ ανθρώπων.

Οι παραπάνω διαδικασίες στηρίζονται στη χρονική εξέλιξη του φαινομένου της διάχυσης. Αντίστοιχα, μπορεί κανείς να μελετήσει τη μόνιμη κατάσταση του φαινομένου αυτού αντί να εκτελεί τις επαναλήψεις διαδοχικά. Προς την κατεύθυνση αυτή, παρουσιάστηκε ο αλγόριθμος Random Walker και μια γενίκευση του (Power Watershed). Είδαμε ότι η χρήση ενός RAG γράφου αντί ενός κανονικού πλέγματος μειώνει δραματικά τους χρόνους εκτέλεσης.

Το δεύτερο κύριο αποτέλεσμα του κεφαλαίου αυτού, συνδυάζει τη γνώση των PDE και τον αλγόριθμο του Τυχαίου Περιπατητή. Πιο συγκεκριμένα, χρησιμοποιήθηκε η κανονικοποιημένη Graph Laplacian ώστε να αναπαριστώνται με μεγαλύτερη ακρίβεια οι σχέσεις μεταξύ των κόμβων ανάλογα με το βαθμό του καθενός από αυτούς. Είδαμε ότι είτε χρησιμοποιήσει κανείς κανονικό πλέγμα είτε RAG γράφο, τα αποτελέσματα (και ειδικά στη δεύτερη περίπτωση) είναι βελτιωμένα.

Τέλος, παρατηρήθηκε ότι πράγματι ο αλγόριθμος του Τυχαίου Περιπατητή είναι ευαίσθητος στην επιλογή των αρχικών σπόρων και ότι η γνώση της παραμέτρου σ για την απόδοση βαρών στις ακμές μεταξύ των κόμβων του γράφου διαδραματίζει καίριο ρόλο στο τελικό αποτέλεσμα της κατάτμησης. Θεωρητικά, με το να εισάγουμε πληροφορία για τους βαθμούς και των 2 κόμβων για το variation επί μιας ακμής, εισάγουμε μια επιπλέον χρήσιμη πληροφορία. Προς την κατεύθυνση αυτή, συνέβαλλε η πειραματική επιβεβαίωση ότι ο Unbiased Random Walker μπορεί να ελαττώσει την αρνητική αυτή επίδραση.

Κεφάλαιο 4

Τεχνικές Εκμάθησης Χωρίς Επίβλεψη (Unsupervised Learning)

4.1 Εισαγωγή στη μέθοδο K-Means

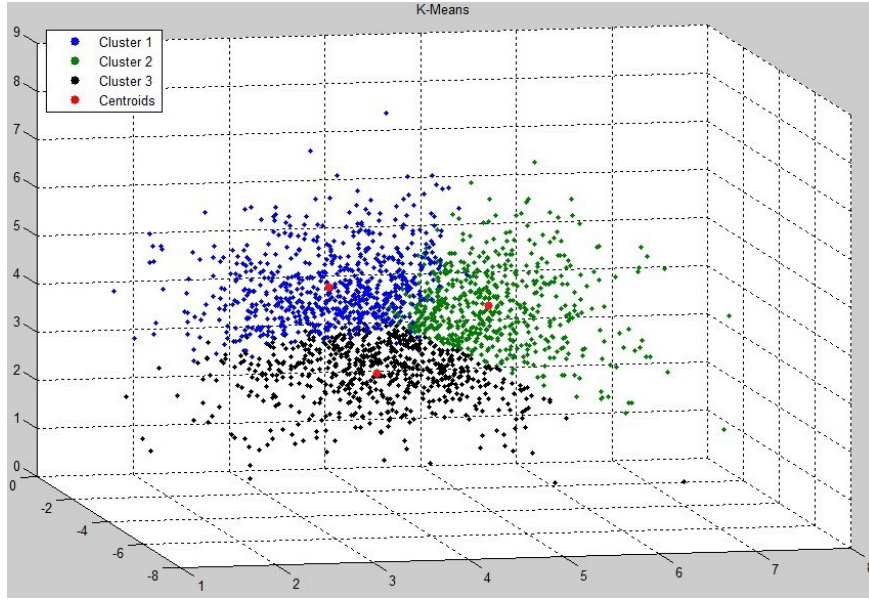
Ο αλγόριθμος K-Means είναι ίσως η πιο διάσημη unsupervised μέθοδος, μέθοδος δηλαδή στην οποία ο χρήστης δεν εισάγει κάποια πληροφορία. Δεδομένου ενός συνόλου παρατηρήσεων $\{x_1, x_2, \dots, x_n\}$ με $x_i \in \mathcal{R}^d \forall i = 1, 2, \dots, n$, βασικός στόχος είναι η διαμέριση των παρατηρήσεων σε k ($k < n$) σύνολα $S = \{S_1, S_2, \dots, S_k\}$ ώστε να ελαχιστοποιείται η ενδοταξική απόσταση από τα κεντροειδή μ_i καθενός cluster. Με άλλα λόγια ζητείται η

ελαχιστοποίηση $\arg \min_S \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$. Όπως σημειώνεται και στο ([15]), ο K-Means

αλγόριθμος μπορεί να γραφεί ως $\max_{G^T G = I} G^T K G$, όπου $K = X^T X$ είναι ένα γραμμικό kernel διάστασης $n \times n$, ο X ένας $d \times n$ πίνακας που περιέχει όλα τα features και ο G ένας $n \times k$ ενδείκτης πίνακας ο οποίος δείχνει την πιθανότητα καθενός από τα n σημεία να ανήκει σε κάθεμία από τις k κλάσεις (0 ή 1). Γενικότερα, για την kernelized εκδοχή του K-Means έχουμε $K = \langle \phi(X), \phi(X) \rangle$. Στη γραμμική περίπτωση, η συνάρτηση $\phi(X) = X$.

Η βασική ιδέα του αλγορίθμου είναι η εξής: δεδομένης μιας αρχικοποίησης των k κεντροειδών, στο πρώτο βήμα ανατίθενται τα σημεία στο κοντινότερο (με την έννοια της ευκλείδειας απόστασης) κεντροειδές που τους αντιστοιχεί (assignment step). Στο δεύτερο βήμα, υπολογίζονται τα νέα κεντροειδή τα οποία προκύπτουν με βάση τα ταξινομημένα ανά κλάση σημεία (update step). Η εκτέλεση των 2 αυτών βημάτων επαναλαμβάνεται εναλλάξ μέχρι να επιτευχθεί σύγκλιση, δηλαδή είτε ολοκληρωθεί ένας προκαθορισμένος αριθμός επαναλήψεων είτε η μεταβολή της θέσης των κεντροειδών μεταξύ δύο διαδοχικών βημάτων είναι κάτω από ένα προεπιλεγμένο κατώφλι.

Το βασικό πλεονεκτήμα του αλγορίθμου είναι η απλή και αποτελεσματική του υλοποίηση. Όμως, δεν υπάρχει εγγύηση ότι η λύση η οποία προκύπτει είναι η βέλτιστη δυνατή παρά μόνο σε τοπικό επίπεδο. Επιπλέον, υπάρχει ευαισθησία στην εκάστοτε αρχικοποίηση, στον αριθμό k των κλάσεων που φάχνουμε και δε λαμβάνεται υπόψη η γεωμετρία του προβλήματος. Για τους λόγους αυτούς, χρησιμοποιείται συνήθως σαν προπαρασκευαστικό στάδιο πιο εξελιγμένων μεθόδων.



Σχήμα 4.1: Ενδεικτικό αποτέλεσμα K-Means, 3 classes.

4.2 Εισαγωγή στη μέθοδο NMF

Όπως σημειώθηκε στο [15], υπάρχει στενή σύνδεση του αλγόριθμου K-Means με μια άλλη πολύ ενδιαφέρουσα μέθοδο clustering, γνωστή με τον όρο Non-Negative Matrix Factorization (NMF). Η βασική ιδέα ([6]) αφορά την επίλυση του προβλήματος

$$\min_{F,G} \|X - F_+ G_+^\top\|^2 \text{ ώστε } X \approx F_+ G_+^\top$$

όπου ο X είναι πάλι $d \times n$, ο F_+ είναι $d \times k$ και ο G_+ είναι $n \times k$ (ο υποδείκτης + δηλώνει ότι ζητάμε αυτοί οι πίνακες να είναι μη αρνητικοί). Είναι γνωστό ότι το παραπάνω πρόβλημα ελαχιστοποίησης δεν είναι κυρτό ταυτόχρονα για τους δύο πίνακες, είναι όμως κυρτό ως προς τον έναν από τους δύο όταν ο άλλος θεωρηθεί γνωστός. Επομένως, δεν εξασφαλίζεται ολικό ελάχιστο παρά τοπικό ελάχιστο. Η εύρεση του ελαχίστου γίνεται εφαρμόζοντας επαναληπτικά (μέχρι να επιτευχθεί σύγκλιση) τις ακόλουθες επαναληπτικές εξισώσεις:

$$G \leftarrow G \odot \frac{F^\top X}{F^\top F G}$$

$$F \leftarrow F \odot \frac{X G^\top}{F G G^\top}$$

όπου το σύμβολο \odot συμβολίζει την πράξη του πολλαπλασιασμού στοιχείο προς στοιχείο μεταξύ 2 πινάκων και το $/$ συμβολίζει την αντίστοιχη διαίρεση. Σημειώνεται ότι για τον κλασικό αλγόριθμο NMF η ελαχιστοποίηση δεν έχει μοναδική λύση καθώς για ένα ζευγάρι λύσεων F, G κάθε άλλο ζευγάρι της μορφής $F D, G D^{-1}$ αποτελεί λύση (για κάθε τυχαίο διαγώνιο πίνακα D), οπότε μπορούμε να πάρουμε μοναδική λύση κανονικοποιώντας στήλες του F να έχουν μήκος 1 και αντίστοιχα τον G ώστε το γινόμενο $F G^\top$ να παραμένει αμετάβλητο:

$$g_{ij} = g_{ij} \sqrt{\sum_i f_{ij}^2} \text{ και } f_{ij} = \frac{f_{ij}}{\sum_i f_{ij}^2}$$

4.2.1 Επεκτάσεις της μέθοδου NMF

Για την αντιμετώπιση παρόμοιων ελαχιστοποιήσεων έχουν προταθεί και αντίστοιχα variants του παραδοσιακού NMF, όπως το Convex NMF, Convex Hull NMF, Semi NMF κλπ ([15]). Το κίνητρο πίσω από τη μελέτη τέτοιων παραγοντοποιήσεων αντί πιο διαδεδομένων όπως η SVD, προκύπτει από το γεγονός ότι σε πολλές εφαρμογές οι πίνακες των χαρακτηριστικών είναι μη αρνητικοί όπως, για παράδειγμα, οι φωτεινότητες των pixel στις εικόνες. Η υπεροχή των μεθόδων NMF έναντι άλλων όπως PCA, SVD έχει βρει έδαφος και σε άλλες εφαρμογές, όπως η ταξινόμηση αρχείων ([22]), αλλά και η αναγνώριση προσώπων ([38]).

4.2.2 Cluster NMF

Μια ενδιαφέρουσα και χρήσιμη παραλλαγή στον αλγόριθμο Convex NMF (όπου θεωρούμε ότι ο πίνακας F μπορεί να γραφεί σα γραμμικός συνδυασμός των δεδομένων στον πίνακα X) είναι η περίπτωση του Cluster NMF. Στην περίπτωση αυτή ερμηνεύουμε τον πίνακα G ως τον πίνακα ο οποίος περιέχει τις ύστερες πιθανότητες να ανήκει το κάθε data point σε ένα cluster. Τότε, μπορούμε να γράψουμε (έστω k ο αριθμός των clusters και n_i ο αριθμός των data points τα οποία ανήκουν στο i cluster) $f_i = X \frac{g_i}{n_i}$ και αν $D = \text{diag}(n_1, n_2, \dots, n_k)$ τότε $F = XGD^{-1}$. Προφανώς, όταν τα δεδομένα στον X είναι μη αρνητικά προκύπτει ότι ο πίνακας F είναι μη αρνητικός, ενώ ο G είναι μη αρνητικός με βάση τον ορισμό του.

Επομένως, ο πίνακας F περιέχει τα κεντροειδή των clusters και η προσέγγιση $X \approx FG^T$ μετατρέπεται στην $X \approx XGD^{-1}G^T$. Ενσωματώνοντας στον G την ποσότητα $D^{-\frac{1}{2}}$ προκύπτει η ελαχιστοποίηση:

$$\min_G \|X - XGG^T\|^2 \quad (4.1)$$

Τέλος, το clustering γίνεται με βάση τον κανόνα

$$\text{label}_i = \arg \max_j G_{i,j} \quad (4.2)$$

καθώς κάθε γραμμή i του πίνακα G εκφράζει το πόσο πιθανό είναι να ανήκει ο κόμβος i στην κλάση j (της αντίστοιχης στήλης j). Επιπρόσθετα, παρατηρούμε ότι η ελαχιστοποίηση που χρησιμοποιεί η μέθοδος είναι η ίδια που προκύπτει από την ανάλυση GLNP (θεωρώντας τον πίνακα X μη αρνητικό):

$$\|X - GG^T X\|^2 = \sum_{ij} (X_{ij} - [GG^T X]_{ij})^2 = \|X^T - X^T GG^T\|^2$$

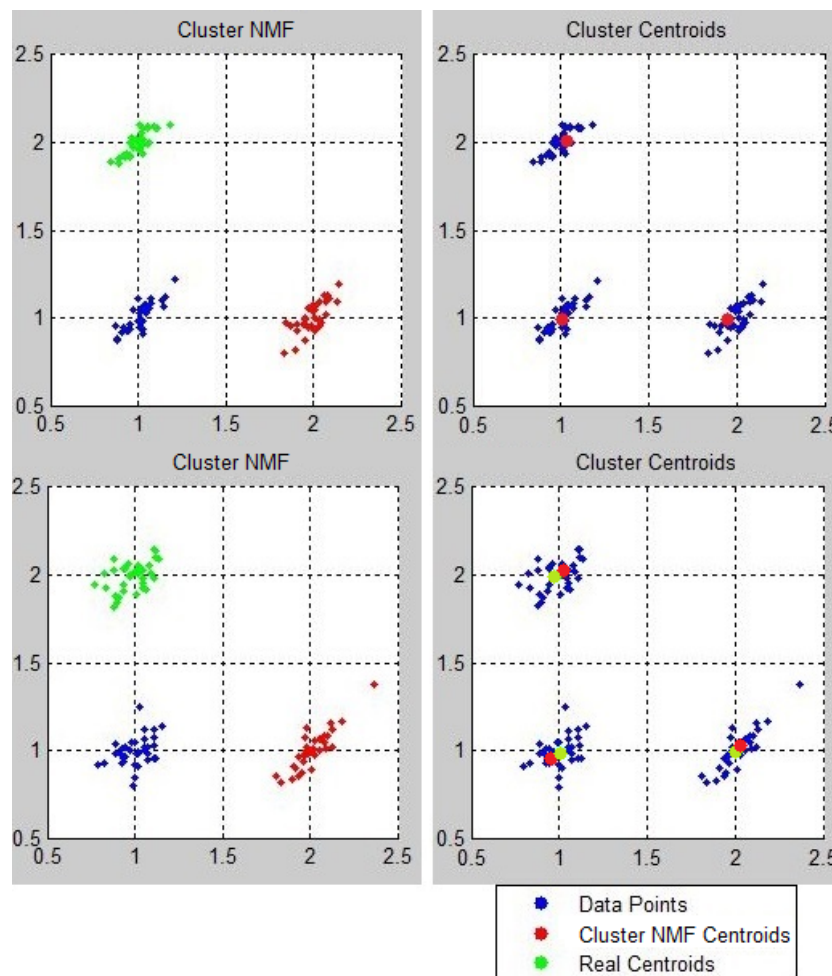
Επίσης, παρατηρούμε ότι (με $\text{Tr}(A)$ συμβολίζουμε το ίχνος του πίνακα A):

$$\begin{aligned} \|X - GG^T X\|^2 &= \sum_{ij} (X_{ij} - [GG^T X]_{ij})^2 = \text{Tr}(X^T X - G^T X^T X G) \\ &\Rightarrow \min_G \|X - XGG^T\|^2 = \min_G \left(\text{Tr}(X^T X) - \text{Tr}(G^T X^T X G) \right) \end{aligned} \quad (4.3)$$

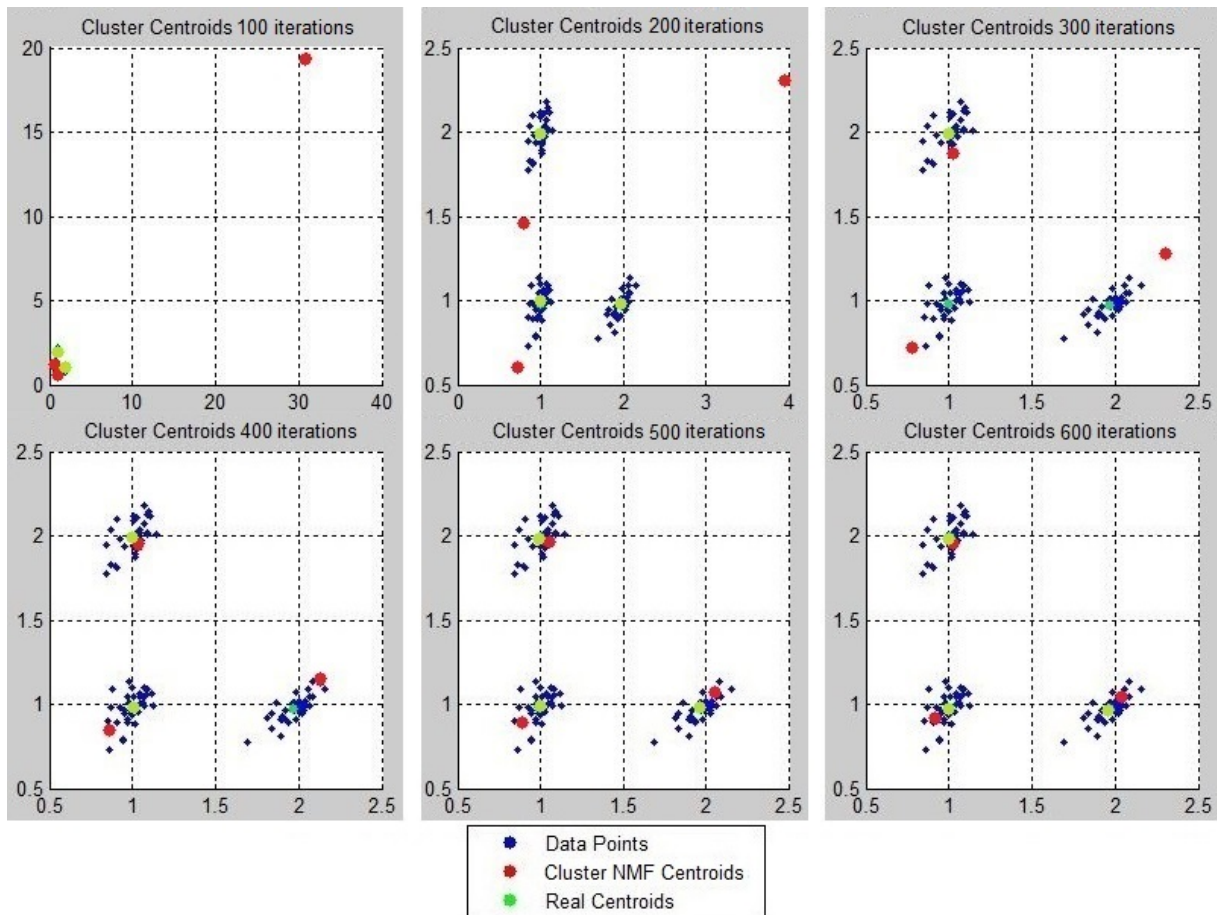
οπότε άμεσα (η ποσότητα $\text{Tr}(X^T X)$ λειτουργεί ως σταθερά στην ελαχιστοποίηση ως προς G , οπότε δεν τη λαμβάνουμε υπόψη) προκύπτει η σύνδεση με τον K-Means αλγόριθμο

υπό τον περιορισμό $G^T G = I$, δηλαδή την απαίτηση ο $G^T G$ να είναι ο μοναδιαίος πίνακας τάξης k . Στην περίπτωση του Cluster NMF, δεν απαιτούμε ο $G^T G$ να είναι μοναδιαίος, δηλαδή ο Cluster NMF αποτελεί μια soft - relaxed εκδοχή του K-Means. Η έννοια soft αναδεικνύει το γεγονός ότι ο πίνακας G δεν είναι ένας πίνακας ο οποίος αποτελείται μόνο από μηδενικά και άσους και μάλιστα έναν άσο σε κάθε γραμμή (ώστε το άθροισμα ανά γραμμές να είναι μονάδα). Αντίθετα, περιέχει θετικές τιμές οι οποίες εκφράζουν ποσοτικά το membership κάθε κόμβου ανά κλάση.

Το γεγονός ότι η ελαχιστοποίηση δεν είναι κυρτό πρόβλημα συνεπάγεται ότι η λύση που θα πάρουμε δεν θα είναι απαραίτητα ολικό ελάχιστο συνεπώς θα υπάρχει εξάρτηση από την αρχικοποίηση του πίνακα G (όπως σημειώνεται στο [3]). Μια πιθανή λύση η οποία αντιμετωπίζει μερικώς την ευαισθησία αυτή είναι να εκμεταλλευτεί κανείς τη σύνδεση μεταξύ Cluster NMF και K-Means οπότε και να αρχικοποιείται ο πίνακας G με βάση τον K-Means αλγόριθμο, όπως περιγράφεται στο ([15]). Από τα παρακάτω διαγράμματα, βλέπουμε ότι όντως τα πραγματικά κεντροειδή των κλάσεων προσεγγίζονται από τα κεντροειδή τα οποία θα δώσει η μέθοδος Cluster NMF. Η προσέγγιση αυτή βελτιώνεται καθώς αυξάνονται οι επαναλήψεις, ώσπου τελικά να φτάσει κανείς στη σύγκλιση της επαναληπτικής μεθόδου. Βέβαια, απαιτείται η πρότερη γνώση του αριθμού των κλάσεων για το clustering.



Σχήμα 4.2: Εκτέλεση Cluster NMF, 3 clusters.



Σχήμα 4.3: Εκτέλεση Cluster NMF, Βελτίωση κεντροειδών ανά iteration, 3 clusters.

4.3 Εφαρμογή του Cluster NMF στην κατάτμηση εικόνας

4.3.1 Αρχική μοντελοποίηση

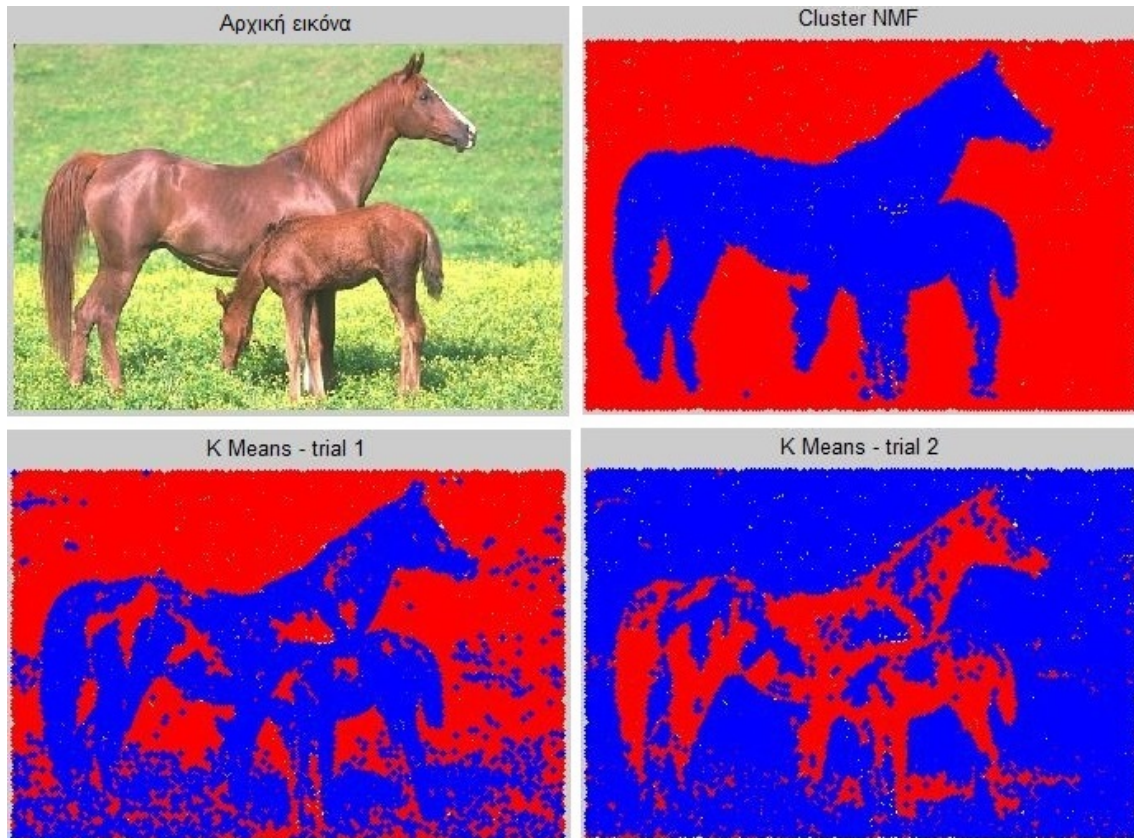
Για τους σκοπούς της εφαρμογής του Cluster NMF, ορίζουμε αρχικά τον πίνακα των χαρακτηριστικών σαν τον πίνακα \mathcal{X} ο οποίος περιέχει αποκλειστικά χρωματική πληροφορία του χρωματικού χώρου RGB. Επομένως, ο πίνακας \mathcal{X} είναι διάστασης $3 \times n$, όπου n είναι ο αριθμός των κόμβων του Region Adjacency γράφου. Μέχρι το σημείο αυτό, δεν εισάγουμε κάποια χωρική πληροφορία, οπότε η τελική κατάτμηση δε θα χρησιμοποιεί τη γεωμετρία του γράφου, παρά μόνο τις μέσες τιμές φωτεινότητας ανά κανάλι του RGB στις περιοχές που έχουν σχηματιστεί σε προηγούμενο στάδιο, μετά από εφαρμογή του watershed μετασχηματισμού.

Οι αρχικές εικόνες για τα πειράματα θα είναι:

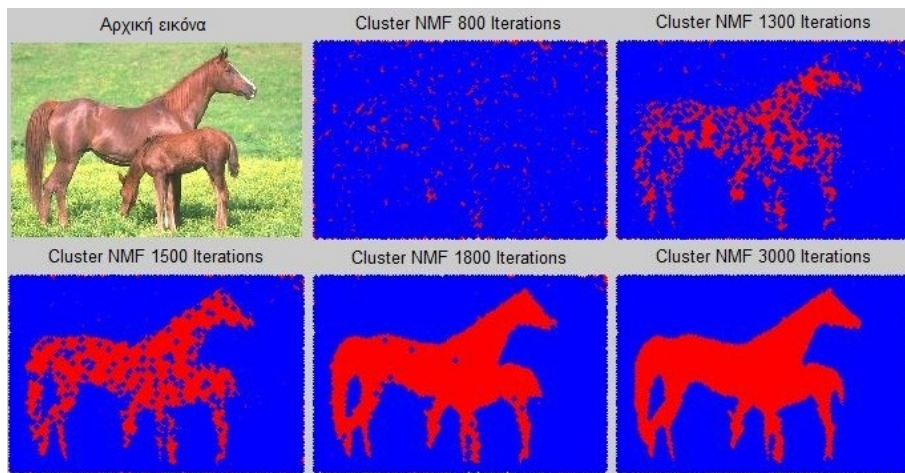


Σχήμα 4.4: Αρχικές Εικόνες, plane, swan, horses.

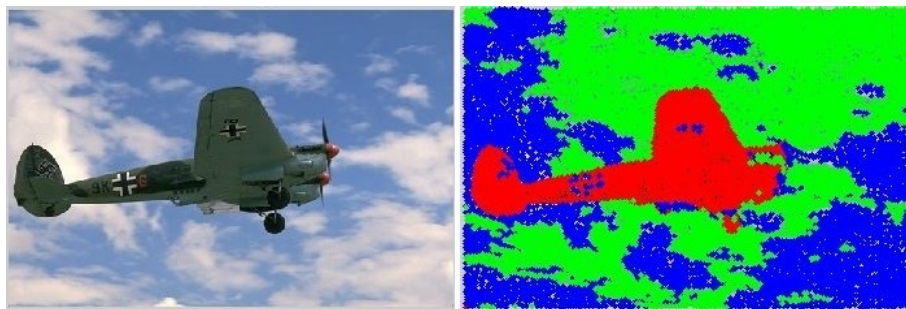
Αρχικά, θα συγκρίνουμε τα αποτελέσματα αν εφαρμόσουμε K-Means μονάχα ή αν εφαρμοστεί Cluster NMF. Από τα παρακάτω αποτελέσματα, παρατηρούμε την αναμενόμενη υπεροχή του Cluster NMF, όπως αυτή έχει καταγραφεί στη βιβλιογραφία για τις τεχνικές NMF εν γένει ([15]). Επίσης, δείχνουμε ότι η τυχαία αρχικοποίηση για τον Cluster NMF απαιτεί πολύ περισσότερες επαναλήψεις για να συγκλίνει, σε σχέση με την αρχικοποίηση με βάση τον K-Means.



Σχήμα 4.5: Cluster NMF vs K-Means, $k = 2$, 500 Iterations για Cluster NMF, horses.



Σχήμα 4.6: Εκτέλεση Cluster NMF με τυχαία αρχικοποίηση, Iterations = [800, 1300, 1500, 1800, 3000], horses.



Σχήμα 4.7: Εκτέλεση Cluster NMF $k = 3$, 500 Iterations για Cluster NMF, plane.



Σχήμα 4.8: Εκτέλεση Cluster NMF $k = 2$, 500 Iterations για Cluster NMF, swan.

Παρατηρώντας τα 2 παραπάνω αποτελέσματα, βλέπουμε ότι η χρωματική πληροφορία δεν είναι πάντα επαρκής για να δώσει ικανοποιητικά αποτελέσματα. Συγκεκριμένα, όπως αναφέρεται και στο [20], υπάρχει υπερκατάτμηση και θορυβώδη όρια μεταξύ των περιοχών (όπως το μπροστινό μέρος του αεροπλάνου και το κάτω μέρος του λαιμού στον κύκνο).

4.3.2 Multiview Learning - συγκρίνοντας με το Cluster NMF

Η εναλλακτική του multiview learning ([20],[21]) αφορά το συνδυασμό χαρακτηριστικών τα οποία ανήκουν σε διαφορετικούς χώρους (views) σε ένα ενιαίο framework ώστε αυτά να μπορούν να περιγράψουν συνολικά τα αντικείμενα. Προς την κατεύθυνση αυτή, αξίζει να αναφερθούμε στις τεχνικές του data integration όπου έχουμε 3 κύριες επιλογές:

1. early integration: απευθείας συνένωση των δεδομένων από όλα τα views (ροές πληροφορίας) σε έναν ενιαίο πίνακα χαρακτηριστικών και έπειτα εκτέλεση του clustering αλγόριθμου
2. intermediate integration: δημιουργία ξεχωριστών πινάκων χαρακτηριστικών και συνδυασμός τους κατάλληλα προτού αυτά διοχετευθούν στον clustering αλγόριθμο
3. late integration: εφαρμογή του clustering αλγόριθμου χωριστά σε κάθε view και έπειτα συνδυασμός των αποτελεσμάτων

Αναφορικά με την πρώτη εκδοχή, έγιναν δοκιμές του κατά πόσον μπορεί η ενιαία περιγραφή χωρικών όρων και όρων φωτεινότητας να συνδυαστεί. Δυστυχώς, σε καμία περίπτωση δεν ήταν εφικτό να πάρουμε ικανοποιητικά αποτελέσματα, γεγονός το οποίο όπως έχουμε ήδη αναφέρει, βρίσκεται σε συμφωνία και με τη βιβλιογραφία και με τη διαίσθηση.

Μια ενδιαφέρουσα προσέγγιση η οποία συνδυάζει τις εκδοχές 2 και 3 έγινε στο [20]. Συγκεκριμένα, προτείνεται μια τροποποιημένη εκδοχή του NMF με την εισαγωγή δύο διαφορετικών πινάκων χαρακτηριστικών X (ροή χρώματος), Y (ροή χωρικής πληροφορίας) οι οποίοι παραγοντοποιούνται ως $X \approx WH$ και $Y \approx VH$ όπου $X \in \mathbf{R}^{m \times n}$, $Y \in \mathbf{R}^{l \times n}$, $W \in \mathbf{R}^{m \times k}$, $V \in \mathbf{R}^{l \times k}$, $H \in \mathbf{R}^{k \times n}$. Στο τελικό στάδιο προτείνεται η επίλυση του προβλήματος:

$$\min_h F(h) \text{ όπου}$$

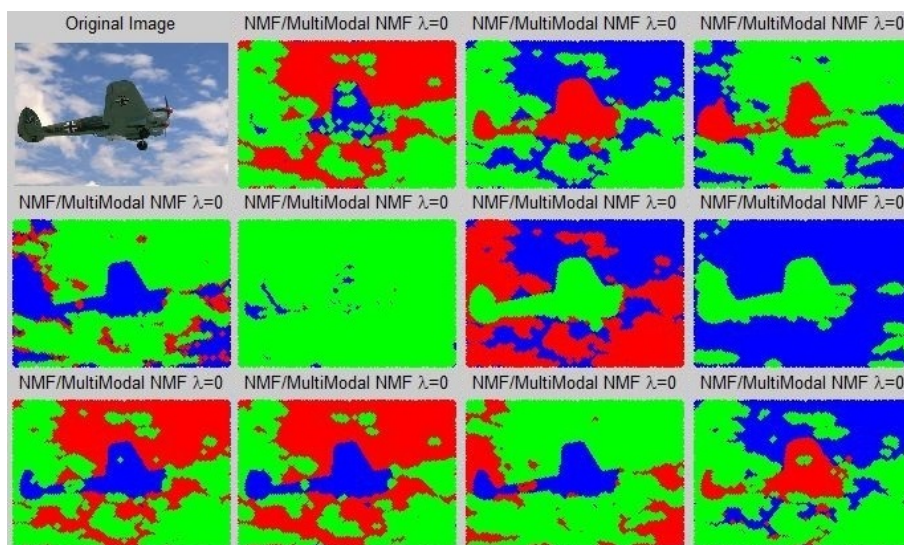
$$\begin{aligned} F(h) &= (1-\lambda) \|x - Wh\|^2 + \lambda \|y - Vh\|^2 = (1-\lambda)(x-Wh)^\top(x-Wh) + \lambda(y-Vh)^\top(y-Vh) \\ &= (1-\lambda)(x^\top x - x^\top Wh - (Wh)^\top x + (Wh)^\top Wh) + \lambda(y^\top y - y^\top Vh - (Vh)^\top y + (Vh)^\top Vh) \\ &= h^\top((1-\lambda)W^\top W + \lambda V^\top V)h - (2(1-\lambda)x^\top W + 2\lambda y^\top V)h + (1-\lambda)x^\top x + \lambda y^\top y \end{aligned}$$

όπου το λ εκφράζει το tradeoff μεταξύ των 2 ροών. Θέτοντας $H(\lambda) = (1-\lambda)W^\top W + \lambda V^\top V$, $f(\lambda)^\top = 2(1-\lambda)x^\top W + 2\lambda y^\top V$ επιλύουμε το πρόβλημα

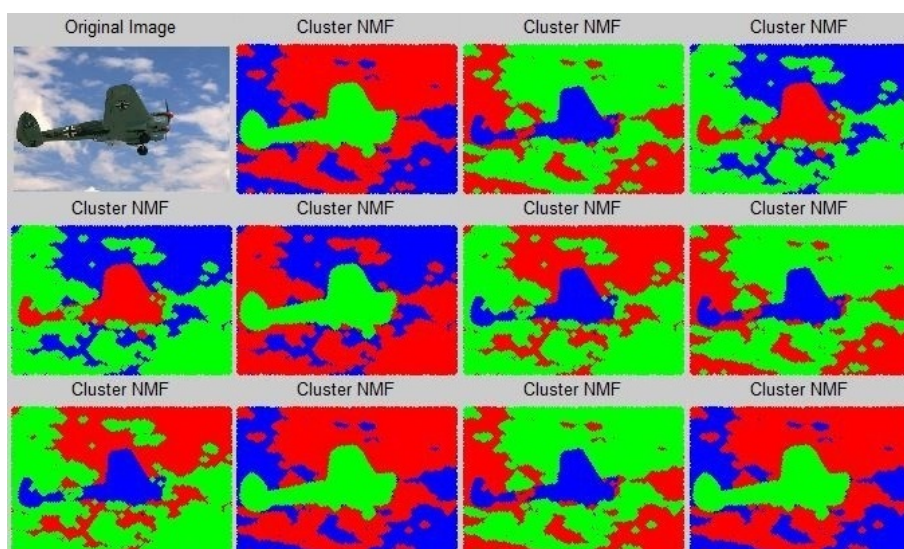
$$\min_h h^\top H(\lambda)h + f(\lambda)^\top h$$

και διαλέγουμε για κάθε data point το cluster c του ως: $c = \arg \max_j h_j$.

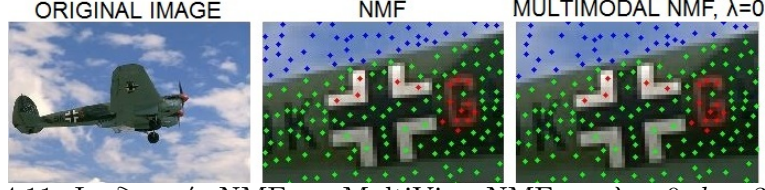
Προφανώς, όταν $\lambda = 0$ έχουμε την περίπτωση όπου χρησιμοποιούμε μόνο το view του color information, εκτελούμε δηλαδή την παραγοντοποίηση NMF αποκλειστικά για τον πίνακα X που περιέχει τις φωτεινότητες των κόμβων. Στη συνέχεια, εκτελούμε 10 πειράματα για NMF και MultiView NMF (όταν $\lambda = 0$) και 10 για τον αλγόριθμο Cluster NMF. Παρατηρούμε ότι στην περίπτωση του NMF υπάρχει πολύ μεγαλύτερη ευαισθησία στις αρχικές συνθήκες. Επίσης, παρατηρούμε (όπως είχε γίνει αντιληπτό και στον Cluster NMF) ότι απαιτείται η εισαγωγή χωρικών όρων για να αποφύγουμε την υπερκατάτμηση. Τέλος, επειδή η επίλυση του MultiView NMF στηρίζεται στον αλγόριθμο NMF, θα επικεντρωθούμε κυρίως στο να επεκτείνουμε τον αλγόριθμο Cluster NMF.



Σχήμα 4.9: Εκτέλεση NMF και MultiView NMF με $\lambda = 0$, $k = 3$, 10 experiments, plane.



Σχήμα 4.10: Εκτέλεση Cluster NMF $k = 3$, 10 experiments, plane.



Σχήμα 4.11: Ισοδυναμία NMF και MultiView NMF για $\lambda = 0$, $k = 3$, plane.

4.4 Κατασκευή - Ενσωμάτωση χωρικού όρου

Όπως παρατηρήθηκε στο [29], η ελαχιστοποίηση μέσω του NMF και των variants αυτού, μπορεί να αποδώσει parts - based αναπαραστάσεις, δίχως να ανακαλύπτει τη γεωμετρική δομή και φύση του προβλήματος. Για το λόγο αυτό, προτάθηκε στο [29] η τροποποίηση των κλασικών multiplicative εξισώσεων του NMF με σκοπό την ελαχιστοποίηση του συναρτησιακού

$$\mathcal{J} = \|X - FG^T\|^2 + \lambda \text{Tr}(G^T \mathcal{L}G)$$

όπου $\mathcal{L} = \mathcal{D} - \mathcal{W}$, $\mathcal{W} = [w_{i,j}]$ $n \times n$ συμμετρικός πίνακας και $\mathcal{D} = [d_{i,i}]$ $n \times n$ διαγώνιος πίνακας με $d_{i,i} = \sum_j w_{i,j}$, όπου

$$w_{i,j} = \begin{cases} 1 & \text{αν } i \sim j, \text{ με βάση τον } k - nn \text{ γράφο} \\ 0 & \text{αλλιώς} \end{cases}$$

Κατ' αναλογία, με σκοπό την ενσωμάτωση του χωρικού όρου στο Cluster NMF framework μπορεί κανείς να ελαχιστοποιήσει την εξής ποσότητα:

$$\mathcal{J} = \|\mathcal{X} - GG^T \mathcal{X}\|^2 + \lambda \text{Tr}(G^T \mathcal{L}G) \quad (4.4)$$

με τον πίνακα \mathcal{W} αυτή τη φορά να είναι:

$$w_{i,j} = \begin{cases} 1 & \text{αν } i \sim j, \text{ με βάση τον RAG} \\ 0 & \text{αλλιώς} \end{cases}$$

Κατ' αναλογία με το [29], θεωρούμε τη συνάρτηση $t_q(\mathbf{x}_i) = g_{i,q}$ η οποία αντιστοιχίζει το σημείο \mathbf{x}_i στον άξονα f_q . Υπενθυμίζουμε ότι στην περίπτωση του Cluster NMF ο πίνακας F είναι διαστάσεων $d \times k$ (d ο αριθμός των features και k ο αριθμός των κλάσεων), έχει σα στήλες του τους άξονες f_q ($q = 1..k$) πάνω στους οποίους προβάλλονται τα δεδομένα και προκύπτει σα γραμμικός συνδυασμός των δεδομένων στον πίνακα \mathcal{X} .

Τελικός στόχος είναι η ελαχιστοποίηση ενός συναρτησιακού ο οποίος θα περιέχει smoothness όρους και η ποσότητα $\|t_q\|_M^2$ μπορεί να χρησιμοποιηθεί ως τέτοια. Υπό την αυστηρή συνεχή έννοια του όρου, θεωρούμε ένα συμπαγές manifold $\mathcal{M} \subset \mathcal{R}^m$ και επιλέγουμε $\|t_q\|_M^2 = \int_{\mathbf{x} \in \mathcal{M}} \|\nabla_{\mathcal{M}} t_q\|^2 dP_X(\mathbf{x})$ με $\nabla_{\mathcal{M}} t_q$ να είναι το gradient του t_q στο manifold \mathcal{M} και τα άκρα ολοκλήρωσης να βρίσκονται επί της κατανομής P_X . Η ποσότητα αυτή υπολογίζεται με προσεγγιστικό τρόπο ([29] και αναφορές εντός αυτού). Επίσης, ορίζεται η ποσότητα REG_q ως η διακριτή προσέγγιση της συνεχούς ποσότητας $\|t_q\|_M^2$ επί του άξονα f_q , δηλαδή:

$$REG_q = \frac{1}{2} \sum_{i,j=1}^n w_{i,j} (t_q(x_i) - t_q(x_j))^2 \quad (4.5)$$

$$= \sum_{i=1}^n t_q^2(x_i) d_{i,i} - \sum_{i,j=1}^n w_{i,j} t_q(x_i) t_q(x_j) \quad (4.6)$$

$$= \sum_{i=1}^n g_{i,q}^2 d_{i,i} - \sum_{i,j=1}^n w_{i,j} g_{i,q} g_{j,q} \quad (4.7)$$

$$= g_q^\top \mathcal{D} g_q - g_q^\top \mathcal{W} g_q \quad (4.8)$$

$$= g_q^\top \mathcal{L} g_q \quad (4.9)$$

όπου g_q είναι η q στήλη του G . Έτσι, η ελαχιστοποίηση της ποσότητας REG_q ($\forall q = 1..k$) συμβάλλει στη χωρική ομοιομορφία της τελικής λύσης. Επιδιώκουμε, επομένως, smooth αποτιμήσεις για την τελική λύση:

$$\mathcal{J} = \|\mathcal{X} - GG^\top \mathcal{X}\|^2 + \lambda Tr(G^\top \mathcal{L} G) = \|\mathcal{X} - GG^\top \mathcal{X}\|^2 + \lambda \sum_{q=1}^k REG_q \quad (4.10)$$

με τον όρο $\lambda \geq 0$ να αυξάνει την επίδραση του smoothness όρου.

4.4.1 Μαθηματική προσέγγιση ελαχιστοποίησης

Με απώτερο σκοπό να ελαχιστοποιήσουμε την ποσότητα

$$\mathcal{J}(G, \lambda) = \|\mathcal{X} - GG^\top \mathcal{X}\|_F^2 + \lambda Tr(G^\top \mathcal{L} G) \quad (4.11)$$

ορίζουμε τη Lagrange συνάρτηση \mathcal{K} του προβλήματος εισάγοντας τον περιορισμό ότι $g_{i,j} \geq 0$ ($G = [g_{i,j}]$) με τους Lagrange πολλαπλασιαστές $\Psi = [\psi_{i,j}]$. Επομένως:

$$\frac{\partial \mathcal{K}}{\partial G} = -2(\mathcal{X} - GG^\top \mathcal{X})\mathcal{X}^\top G - 2\mathcal{X}(\mathcal{X}^\top - \mathcal{X}^\top GG^\top)G + \lambda(\mathcal{L} + \mathcal{L}^\top)G + \Psi \quad (4.12)$$

$$\stackrel{\mathcal{L}=\mathcal{L}^\top}{=} -2\mathcal{X}\mathcal{X}^\top G + 2GG^\top \mathcal{X}\mathcal{X}^\top G - 2\mathcal{X}\mathcal{X}^\top G + 2\mathcal{X}\mathcal{X}^\top GG^\top G + 2\lambda \mathcal{L} G + \Psi \quad (4.13)$$

$$= -4\mathcal{X}\mathcal{X}^\top G + 2GG^\top \mathcal{X}\mathcal{X}^\top G + 2\mathcal{X}\mathcal{X}^\top GG^\top G + 2\lambda(\mathcal{D} - \mathcal{W})G + \Psi \quad (4.14)$$

Οι συνθήκες KKT δίνουν $g_{i,j}\psi_{i,j} = 0$ οπότε κατ' αναλογία με τα [29], [3] σχηματίζουμε την εξής πολλαπλασιαστική ανανέωση των $g_{i,j}$:

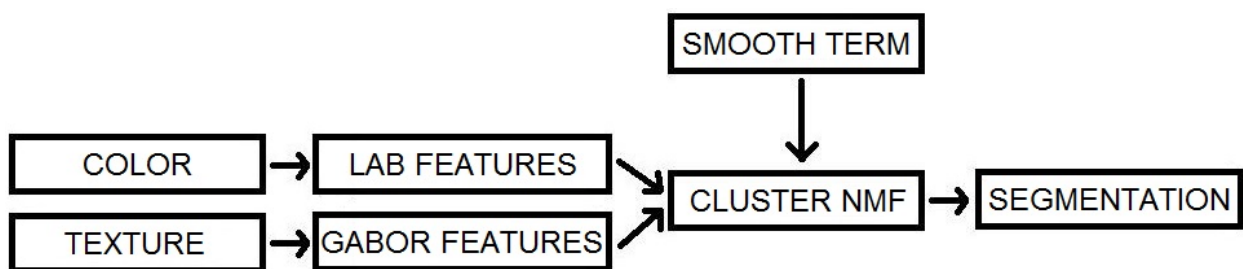
$$g_{i,j} \leftarrow g_{i,j} \sqrt{\frac{[2\mathcal{X}\mathcal{X}^\top G + \lambda \mathcal{W} G]_{i,j}}{[GG^\top \mathcal{X}\mathcal{X}^\top G + \mathcal{X}\mathcal{X}^\top GG^\top G + \lambda \mathcal{D} G]_{i,j}}} \quad (4.15)$$

Σημειώνουμε ότι με την παραπάνω ανάλυση, εισάγουμε πλέον χωρική πληροφορία στην ελαχιστοποίηση μέσω του όρου λ . Μεγάλες τιμές του όρου λ αυξάνουν τη βαρύτητα του χωρικού όρου, ενώ αντίστοιχα για $\lambda \rightarrow 0$ προκύπτουν λύσεις οι οποίες αντιστοιχούν σχεδόν αποκλειστικά στη χρωματική πληροφορία. Η σύγκλιση της (4.15) εξασφαλίζεται από την αναλογία της με τις αντίστοιχες ανανεώσεις των [29], [3] και προέκυψε και πειραματικά.

4.4.2 Περιγραφή της βελτιωμένης μεθόδου

Με βάση την παραπάνω βελτιωμένη ελαχιστοποίηση μέσω του Cluster NMF, μπορούμε να καταστρώσουμε μια αλγοριθμική διαδικασία με στόχο την κατάτμηση αντικειμένων πάνω σε γράφους. Σε πρώτο στάδιο θα επικεντρωθούμε σε 2 περιπτώσεις. Η πρώτη αφορά την κατάτμηση με βάση το χρώμα (οπότε και θα χρησιμοποιήσουμε χαρακτηριστικά χρώματος) και η δεύτερη αφορά την κατάτμηση υφής (texture segmentation). Η βασική μέθοδος αποτελείται από τα εξής βήματα:

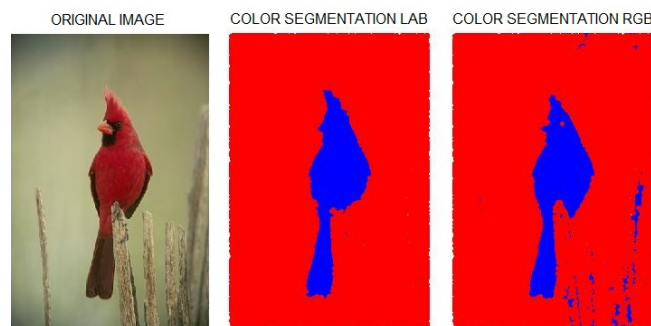
1. εξαγωγή χαρακτηριστικών (χρώματος και υφής)
2. ελαχιστοποίηση της (4.4) με Cluster NMF
3. απόδοση ετικέτας στον i κόμβο με βάση την (4.2) ($\forall i = 1, 2 \dots n$)



Σχήμα 4.12: Βήματα της βελτιωμένης μεθόδου.

4.4.3 Εξαγωγή χαρακτηριστικών χρώματος

Η καλή και ακριβής εξαγωγή χαρακτηριστικών είναι το πρώτο αλλά και ένα πολύ σημαντικό στάδιο της μεθόδου. Στην περίπτωση της κατηγοριοποίησης μιας εικόνας με βάση το χρώμα, κάνουμε μετατροπή του χρωματικού χώρου από RGB σε Lab (κάνουμε και αντίστοιχα ένα rescaling όταν υπάρχουν αρνητικές τιμές στα χαρακτηριστικά). Με τη βοήθεια αυτής της μετατροπής προσπαθεί κανείς να εξαλείψει τις εξαρτήσεις από τη φωτεινότητα (intensity). Στη συνέχεια, από τον πίνακα χαρακτηριστικών X υπολογίζουμε για κάθε κόμβο τις μέσες τιμές του εκάστοτε καναλιού. Μετά τη μετατροπή αυτή, διατηρούνται μόνο τα κανάλια a, b και ο χώρος των χαρακτηριστικών είναι $2 \times n$, όπου n ο αριθμός των κόμβων.



Σχήμα 4.13: Παράδειγμα σύγκρισης RGB με LAB, bird.

4.4.4 Εξαγωγή χαρακτηριστικών υφής - Gabor Filterbank

Για την εξαγωγή των χαρακτηριστικών της υφής, θα χρησιμοποιηθούν τα διαδεδομένα φίλτρα Gabor. Σύμφωνα με τη βασική θεωρία των φίλτρων Gabor, τα ζωνοπερατά αυτά φίλτρα έχει αποδειχθεί ότι μιμούνται τον ανθρώπινο τρόπο ανάλυσης μιας εικόνας, οπότε και χρησιμοποιούνται στην εξαγωγή χαρακτηριστικών υφής. Συγκεκριμένα, ένα φίλτρο Gabor αποτελείται από μια Gaussian συνάρτηση διαμορφωμένη από ένα μιγαδικό επίπεδο κύμα κεντρικής συχνότητας f_u και προσανατολισμού θ_v ([33]).

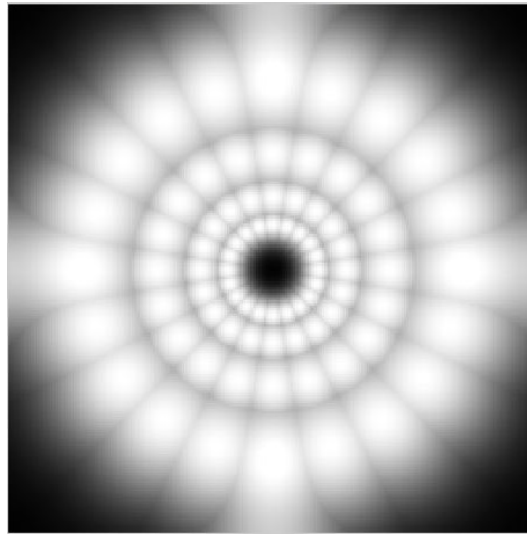
$$\psi_{u,v}(x, y) = \frac{f_u^2}{\pi\gamma\eta} e^{-\left(\frac{f_u^2}{\gamma^2}x'^2 + \frac{f_u^2}{\eta^2}y'^2\right)} e^{j2\pi f_u x'} \quad (4.16)$$

όπου

$$x' = x \cos \theta_v + y \sin \theta_v \quad (4.17)$$

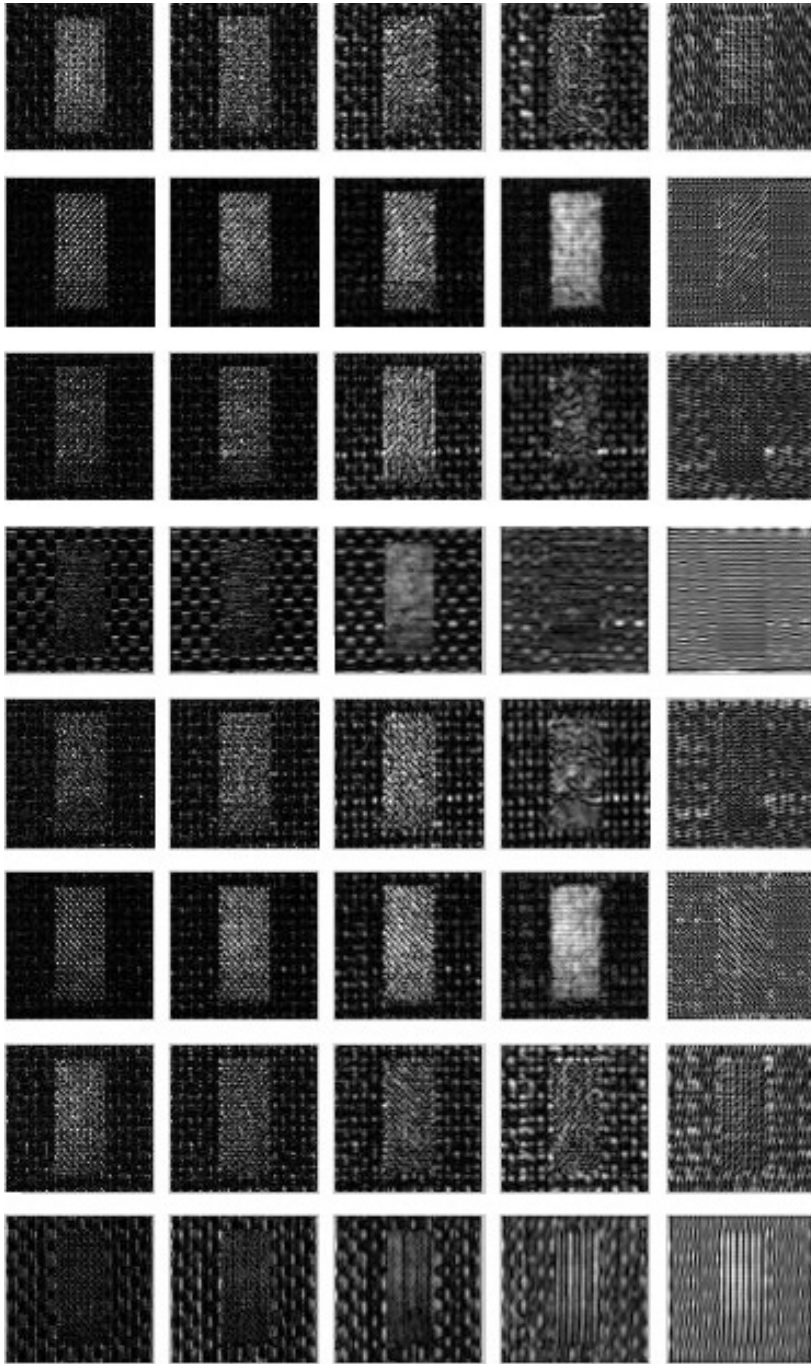
$$y' = -x \sin \theta_v + y \cos \theta_v \quad (4.18)$$

Με απώτερο σκοπό την πολυκλιμακωτή ανάλυση μιας εικόνας, έχει προταθεί η εφαρμογή συστοιχίας τέτοιων φίλτρων. Επομένως, χρησιμοποιούνται s κλίμακες και d προσανατολισμοί ανά κλίμακα ώστε $u \in [1, s], v \in [0, d - 1]$. Θεωρώντας f_{max} τη μέγιστη συχνότητα παίρνουμε ότι: $\theta_v = \frac{v\pi}{d}, f_u = 2^{-\frac{u}{2}} f_{max}$. Ο συνολικός αριθμός των φίλτρων είναι $s \times d$ και για κάθε ένα από αυτά παίρνουμε μια φιλτραρισμένη εκδοχή της αρχικής εικόνας, κρατώντας μόνο τα πλάτη του προκύπτοντος αποτελέσματος.



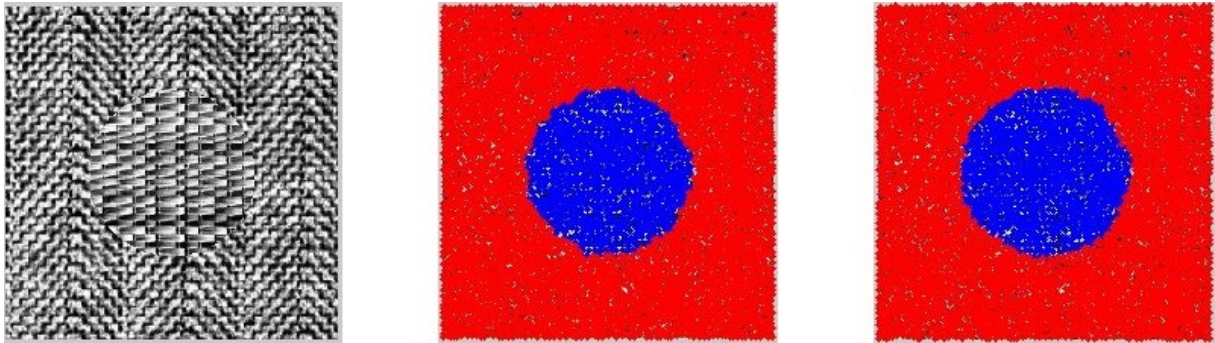
Σχήμα 4.14: Παράδειγμα gabor filterbank, 4 scales and 10 orientations, με τη βοήθεια της υλοποίησης στο [32].

Πιο συγκεκριμένα, στα παρακάτω πειράματα χρησιμοποιήθηκαν 5 κλίμακες και 8 προσανατολισμοί, δηλαδή ο πίνακας \mathcal{X} είναι $40 \times n$, όπου σε κάθε στήλη j περιέχεται η μέση τιμή της απόκρισης στο αντίστοιχο φίλτρο για κάθε κλίμακα και προσανατολισμό στην περιοχή j .

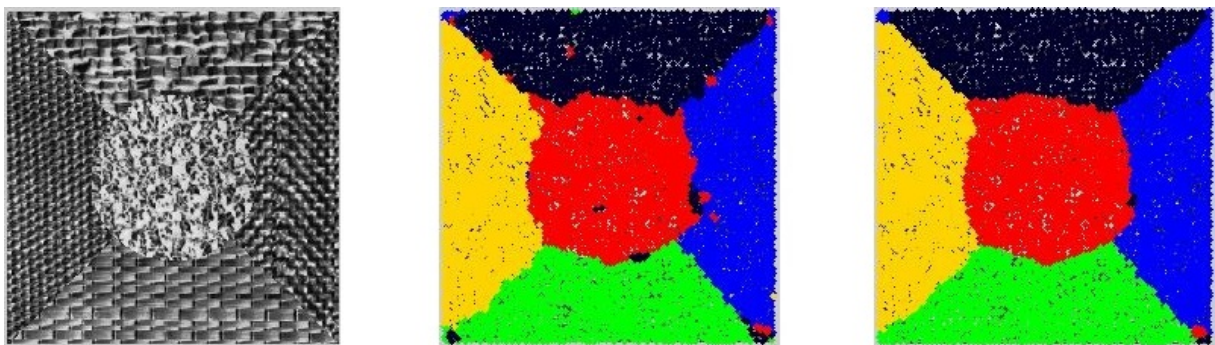


Σχήμα 4.15: Πλάτη φιλτραρισμένων εικόνων, Brodatz 6.

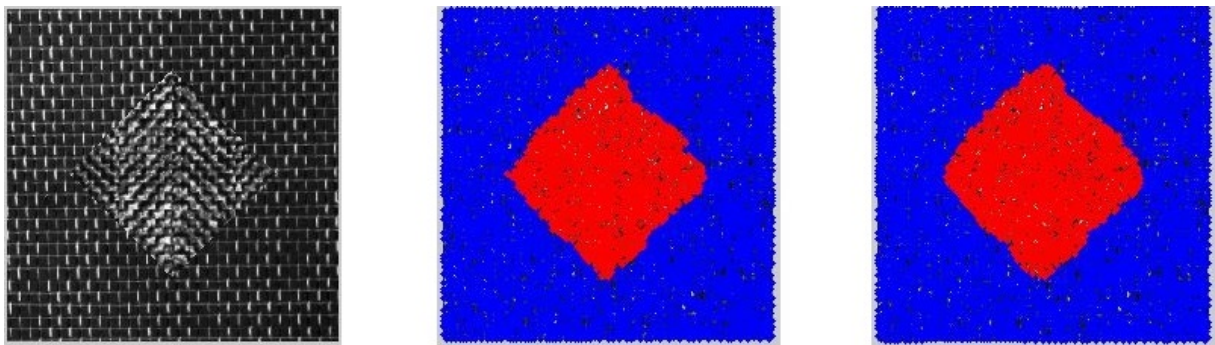
4.4.5 Αποτελέσματα Κατηγοριοποίησης Υφής



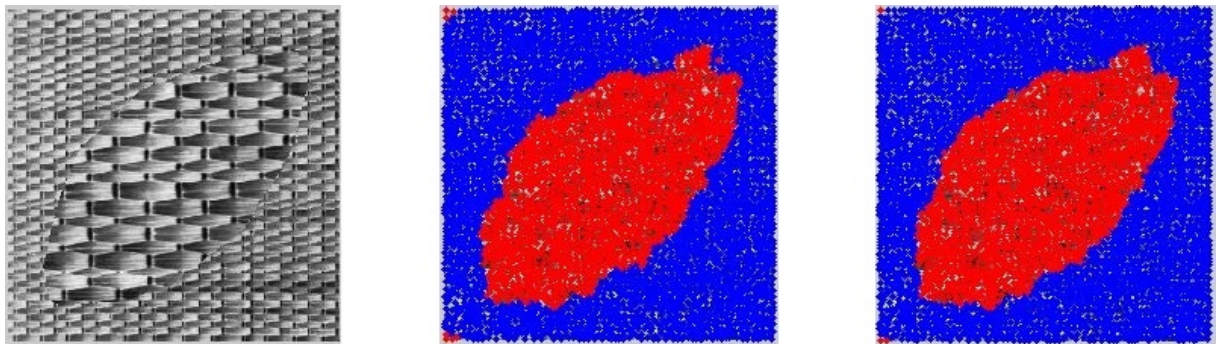
Σχήμα 4.16: Πείραμα με 2 classes, 30 iterations για Cluster NMF, $\lambda = [0, 5]$, Brodatz 1.



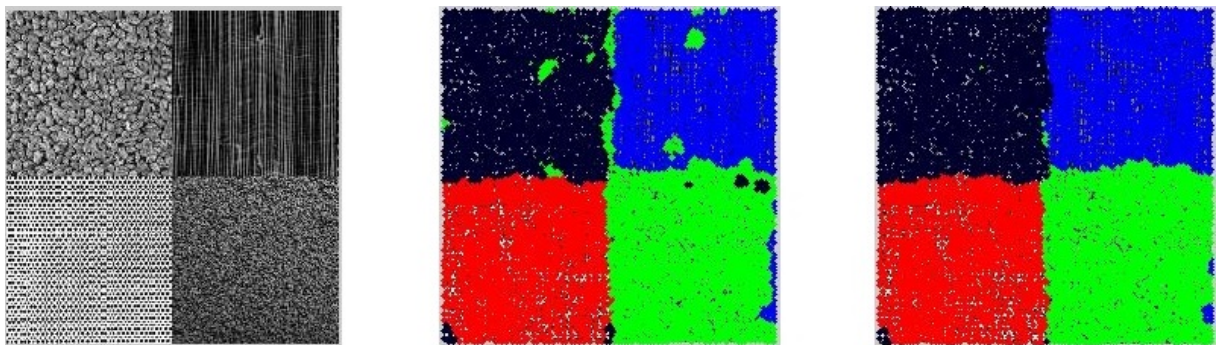
Σχήμα 4.17: Πείραμα με 5 classes, 30 iterations για Cluster NMF, $\lambda = [0, 5]$, Brodatz 2.



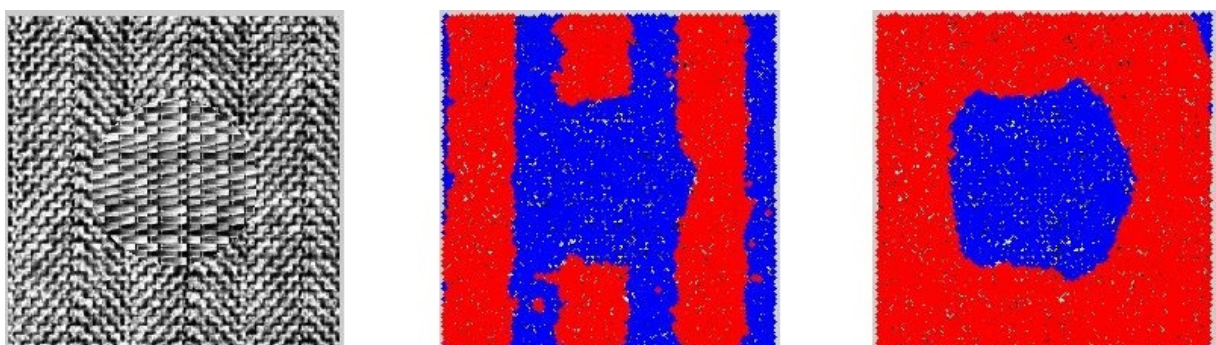
Σχήμα 4.18: Πείραμα με 2 classes, 30 iterations για Cluster NMF, $\lambda = [0, 5]$, Brodatz 3.



Σχήμα 4.19: Πείραμα με 2 classes, 30 iterations για Cluster NMF, $\lambda = [0, 5]$, Brodatz 4.



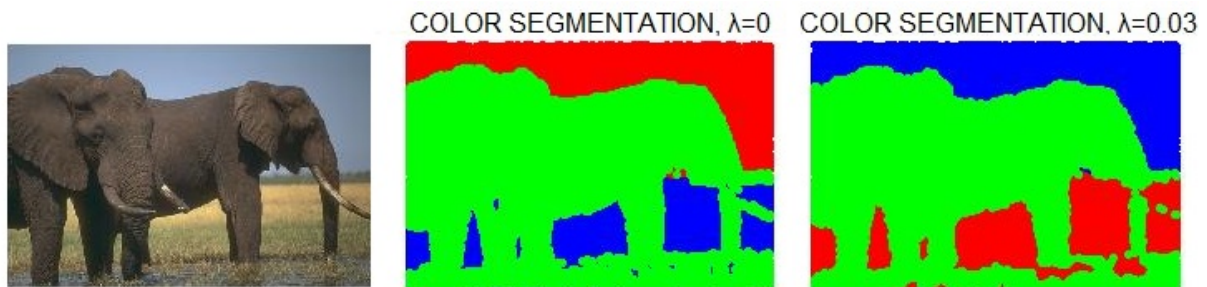
Σχήμα 4.20: Πείραμα με 4 classes, 30 iterations για Cluster NMF, $\lambda = [0, 5]$, Brodatz 5.



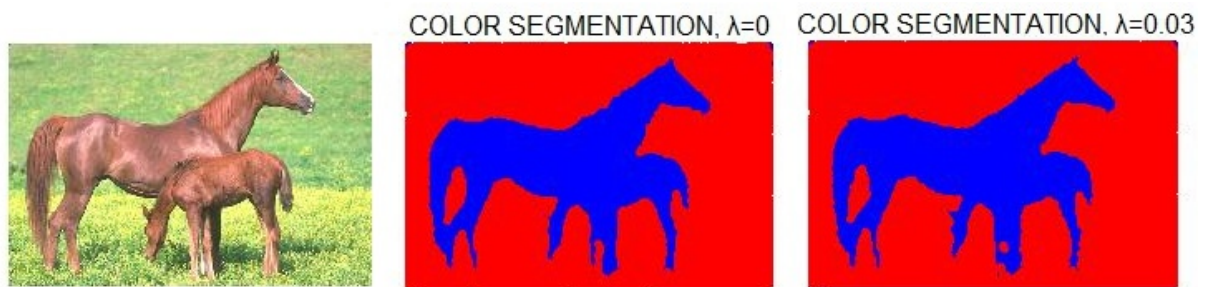
Σχήμα 4.21: Πείραμα με 2 classes, 800 iterations για Cluster NMF, bad result/initialization with spatial improvement, $\lambda = [0, 5]$, Brodatz 1.

4.4.6 Αποτελέσματα Κατηγοριοποίησης Χρώματος

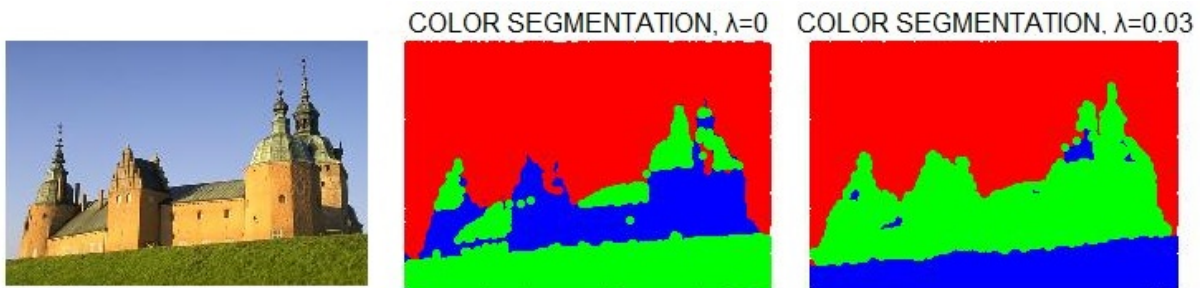
Από τις παρακάτω εικόνες παρατηρούμε το πως επιδρά η παράμετρος λ . Για μια σχετικά μικρή τιμή της παραμέτρου (εμπειρικά επιλέχθηκε 0.03), οι κόμβοι οι οποίοι ταξινομούνται σε κάποια κλάση υποχρεούν σε ένα μέρος και τους γείτονες τους να αποκτούν την ίδια ετικέτα ταξινόμησης με αυτούς. Έτσι, μικρές περιοχές εξαφανίζονται, ενώ κάποιες μεγαλύτερες συρρικνώνονται. Προφανώς, αν χρησιμοποιηθεί πολύ μεγάλη τιμή για την παράμετρο λ , τότε η χρωματική πληροφορία παύει να επιδρά στην ταξινόμηση και άρα στο graph segmetantion.



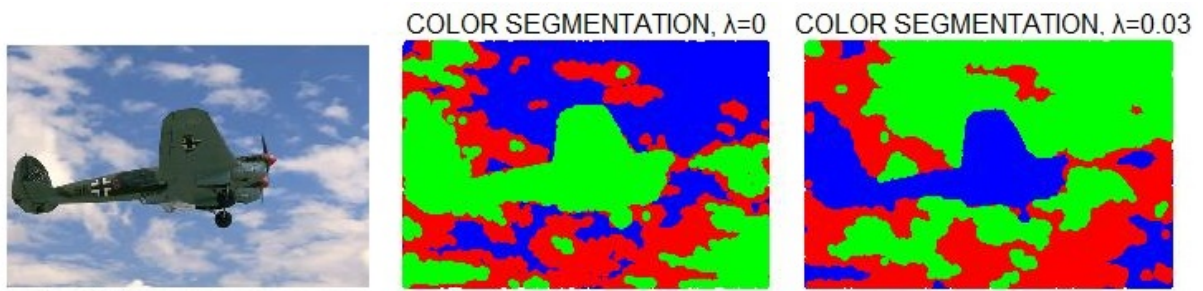
Σχήμα 4.22: Πείραμα με 3 classes, 250 iterations για Cluster NMF, $\lambda = [0, 0.03]$, elephant.



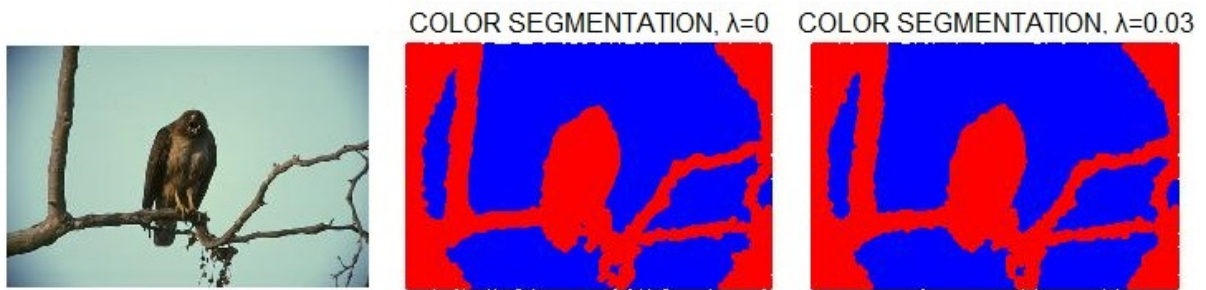
Σχήμα 4.23: Πείραμα με 2 classes, 250 iterations για Cluster NMF, $\lambda = [0, 0.03]$, horses.



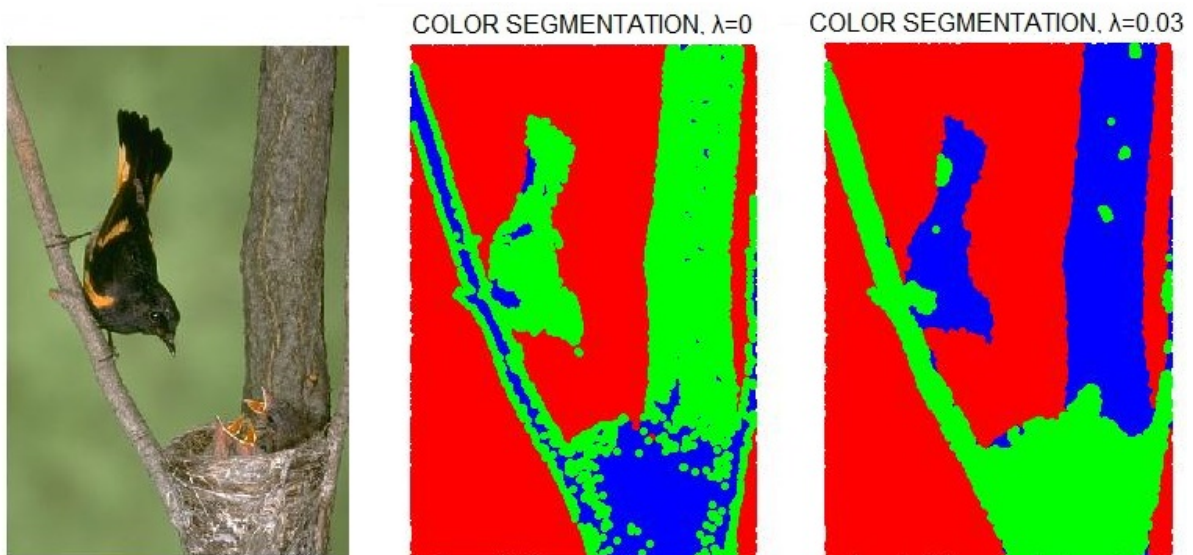
Σχήμα 4.24: Πείραμα με 3 classes, 250 iterations για Cluster NMF, $\lambda = [0, 0.03]$, castle.



Σχήμα 4.25: Πείραμα με 3 classes, 250 iterations για Cluster NMF, $\lambda = [0, 0.03]$, plane.



Σχήμα 4.26: Πείραμα με 2 classes, 250 iterations για Cluster NMF, $\lambda = [0, 0.03]$, eagle.

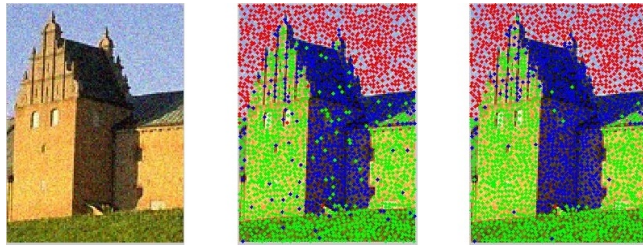


Σχήμα 4.27: Πείραμα με 3 classes, 250 iterations για Cluster NMF, $\lambda = [0, 0.03]$, small bird.

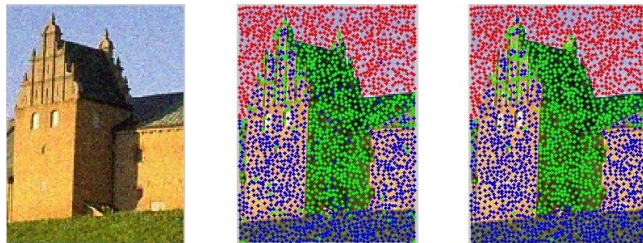
4.4.7 Συνδυάζοντας την ελαχιστοποίηση με το Regularization Framework

Στο κεφάλαιο 2, είδαμε ότι η έννοια του regularization απλοποιεί την αρχική εικόνα και μπορεί να αντιμετωπίσει μερικώς το θόρυβο. Για το λόγο αυτό, μπορεί κανείς να ενσωματώσει λίγες επαναλήψεις regularization στην αρχική εικόνα προτού εφαρμόσει την ελαχιστοποίηση με τη χρήση Cluster NMF. Προφανώς, μπορούμε να θεωρήσουμε εκ νέου το διαχωρισμό στη χρήση προκαθορισμένου σ_{kernel} είτε σταθερού είτε προσδιορισμένου επί της εικόνας ή στη χρήση του adjacency regularization.

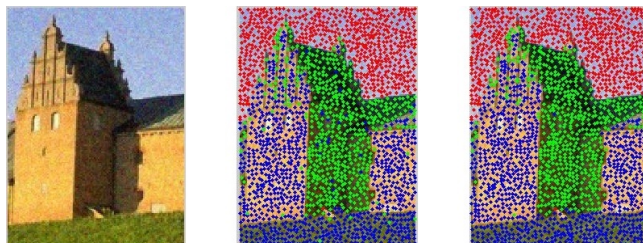
Παρατηρώντας τα ακόλουθα διαγράμματα, βλέπουμε ότι όντως μπορεί να βοηθηθεί η ελαχιστοποίηση ώστε να αντιμετωπίσει το θόρυβο, ειδικά για $\lambda_{spatial} = 0$, δηλαδή όταν δεν εισάγουμε τον όρο χωρικού smoothness στο Cluster NMF. Όταν $\lambda_{spatial} > 0$ τότε φαίνεται ότι η επίδραση του adjacency regularization δεν μας ωφελεί αναγκαστικά. Γενικότερα, καταλήγουμε στο ότι αν εφαρμόσουμε αρκετά ήπιο regularization (όπως για παράδειγμα $\lambda_{reg} = 15, 20$) δίχως τη γνώση του σ_{kernel} (η οποία συνήθως δεν είναι διαθέσιμη), αντισταθμίζουμε καλύτερα την ισορροπία που εισάγει το $\lambda_{spatial} \geq 0$ υπό την ύπαρξη θορύβου.



Σχήμα 4.28: Πείραμα με 3 classes, $NMF_{iter} = 500$, $\lambda_{spatial} = [0, 0.01]$, $\nexists \lambda_{reg}$, 2088 regions, castle.



Σχήμα 4.29: Πείραμα με 3 classes, $NMF_{iter} = 500$, $\lambda_{spatial} = [0, 0.01]$, $REG_{iter} = 100$, $\sigma_{kernel} = 90$, $\lambda_{reg} = 0.01$, 2041 regions, castle.



Σχήμα 4.30: Πείραμα με 3 classes, $NMF_{iter} = 500$, $\lambda_{spatial} = [0, 0.01]$, $REG_{iter} = 100$, $\nexists \sigma_{kernel}$, $\lambda_{reg} = 15$, 2008 regions, castle.

4.5 Consensus Learning

4.5.1 Εισαγωγή

Όπως είδαμε στις προηγούμενες ενότητες, αν θέλει κανείς να εξασφαλίζει βελτιωμένα αποτελέσματα κατάτμησης, απαιτείται ο συνδυασμός ροών πληροφορίας. Στην πιο συνηθισμένη περίπτωση, τέτοιες ροές πληροφορίας είναι το χρώμα και η υφή. Επομένως, το επόμενο βήμα στη βελτίωση του συστήματος κατάτμησης είναι η εισαγωγή τέτοιων μεθόδων οι οποίες να συνδυάζουν διαφορετικές ροές πληροφορίας. Μια τέτοια μέθοδος είναι το Consensus Learning ([31]).

4.5.2 Μαθηματική διατύπωση του Consensus Learning

Έστω μια εικόνα μεγέθους $p \times q$ pixel και το σύνολο $\mathcal{X} = \{x_1, x_2, \dots, x_{pq}\}$. Έστω ακόμα ότι μας έχουν δοθεί T διαμερίσεις $\{\mathcal{S}^1, \mathcal{S}^2, \dots, \mathcal{S}^T\}$, στις οποίες έχουμε και τα αντίστοιχα clusters, δηλαδή $\forall t = 1 \dots T$ έχουμε τα \mathcal{K}_t clusters: $\{\mathcal{S}_1^t, \mathcal{S}_2^t, \dots, \mathcal{S}_{\mathcal{K}_t}^t\}$. Σύμφωνα με το [31], ορίζουμε την απόσταση μεταξύ 2 διαμερίσεων ως

$$\Delta(\mathcal{S}^1, \mathcal{S}^2) = \sum_{i=1}^{pq} \sum_{j=1}^{pq} \delta_{i,j}(\mathcal{S}^1, \mathcal{S}^2)$$

$$\text{όπου } \delta_{i,j} = \begin{cases} 1 & \text{αν } i, j \in \mathcal{S}_k^1 \text{ και } i, j \notin \mathcal{S}_k^2 \text{ (σε κάποιο cluster } k) \\ 1 & \text{αν } i, j \in \mathcal{S}_k^2 \text{ και } i, j \notin \mathcal{S}_k^1 \text{ (σε κάποιο cluster } k) \\ 0 & \text{αλλιώς} \end{cases}$$

Αν αναπαριστήσουμε κάθε διαμέριση \mathcal{S}^t με έναν πίνακα $\mathcal{M} = [m_{i,j}]$, διάστασης $pq \times pq$

$$m_{i,j}(\mathcal{S}^t) = \begin{cases} 1 & \text{αν } i, j \in \mathcal{S}_k^t \text{ (σε κάποιο cluster } k) \\ 0 & \text{αλλιώς} \end{cases}$$

$$\text{τότε } \delta_{i,j} = (m_{i,j}(\mathcal{S}^1) - m_{i,j}(\mathcal{S}^2))^2$$

Η βασική ιδέα είναι η κατασκευή μιας βέλτιστης διαμερίσης \mathcal{S}^* η οποία συνδυάζει με το βέλτιστο τρόπο τις αρχικές διαμερίσεις τις οποίες έχουμε στη διάθεση μας. Η διαμέριση αυτή προκύπτει ελαχιστοποιώντας ποσότητες όπως:

$$\min_{\mathcal{S}^*} \frac{1}{T} \sum_{t=1}^T \Delta(\mathcal{S}^t, \mathcal{S}^*) = \min_{\mathcal{S}^*} \frac{1}{T} \sum_{t=1}^T \sum_{i,j=1}^{pq} (m_{i,j}(\mathcal{S}^t) - m_{i,j}(\mathcal{S}^*))^2$$

Προς τούτο, ορίζουμε τις ποσότητες:

$$\tilde{m}_{i,j} = \frac{1}{T} \sum_{t=1}^T m_{i,j}(\mathcal{S}^t), \quad \Delta \mathcal{M}^2 = \frac{1}{T} \sum_{t=1}^T \sum_{i,j=1}^{pq} (m_{i,j}(\mathcal{S}^t) - \tilde{m}_{i,j})^2 \text{ (σταθερά) και}$$

$$\mathcal{U} = [u_{i,j}] = [m_{i,j}(\mathcal{S}^*)], \quad \tilde{\mathcal{M}} = [\tilde{m}_{i,j}]$$

οπότε ελαχιστοποιούμε αντίστοιχα:

$$\begin{aligned} \min_{\mathcal{S}^*} \frac{1}{T} \sum_{t=1}^T \sum_{i,j=1}^{pq} (m_{i,j}(\mathcal{S}^t) - m_{i,j}(\mathcal{S}^*))^2 &= \min_{\mathcal{S}^*} \frac{1}{T} \sum_{t=1}^T \sum_{i,j=1}^{pq} (m_{i,j}(\mathcal{S}^t) - \tilde{m}_{i,j} + \tilde{m}_{i,j} - m_{i,j}(\mathcal{S}^*))^2 \\ &= \min_{\mathcal{U}} \Delta M^2 + \sum_{i,j=1}^{pq} (\tilde{m}_{i,j} - u_{i,j})^2 \end{aligned}$$

επειδή η ποσότητα ΔM^2 είναι σταθερή ανεξάρτητη της ελαχιστοποίησης και δείχνει την απόκλιση μεταξύ των επιμέρους δεδομένων καταταμήσεων, τελικά παίρνουμε:

$$\min_{\mathcal{U}} \left\| \tilde{\mathcal{M}} - \mathcal{U} \right\|^2$$

Υπάρχουν κάποιοι περιορισμοί στον πίνακα \mathcal{U} οι οποίοι δυσχεραίνουν την επίλυση του παραπάνω προβλήματος. Έστω 3 κόμβοι i, j και l . Το να ανήκει κάποιος κόμβος σε ένα cluster είναι μια μεταβατική σχέση. Δηλαδή, έστω 3 cluster $\mathcal{A}, \mathcal{B}, \mathcal{C}$. Παρατηρούμε τις εξής περιπτώσεις:

$$\text{αν } i, j \in \mathcal{A} \text{ και } j, l \in \mathcal{A} \Rightarrow i, l \in \mathcal{A}, \text{ δηλαδή } u_{i,j} = u_{j,l} = u_{i,l} = 1$$

$$\text{αν } i, j \notin \mathcal{A} \text{ και } j, l \in \mathcal{A} \Rightarrow i, l \notin \mathcal{A}, \text{ δηλαδή } u_{i,j} = 0, u_{j,l} = 1, u_{i,l} = 0$$

$$\text{αν } i \in \mathcal{A} \text{ και } j \in \mathcal{B} \Rightarrow l \in \mathcal{A} \text{ ή } l \in \mathcal{B} \text{ ή } l \in \mathcal{C}, \text{ δηλαδή } \begin{cases} u_{i,j} = 0, u_{j,l} = 0, u_{i,l} = 1 \\ u_{i,j} = 0, u_{j,l} = 1, u_{i,l} = 0 \\ u_{i,j} = 0, u_{j,l} = 0, u_{i,l} = 0 \end{cases}$$

Οι παραπάνω περιπτώσεις κωδικοποιούνται στις ακόλουθες εξισώσεις:

$$u_{i,j} + u_{j,l} - u_{i,l} \leq 1 \quad (4.19)$$

$$u_{i,j} - u_{j,l} - u_{i,l} \leq 1 \quad (4.20)$$

$$-u_{i,j} + u_{j,l} + u_{i,l} \leq 1 \quad (4.21)$$

Παρατηρείται επομένως, ότι σε κάθε κόμβο υπάρχουν 3 τέτοιοι περιορισμοί. Η επίλυση ενός τέτοιου προβλήματος βελτιστοποίησης είναι πολύ δύσκολη και λύνεται με διαφορετική προσέγγιση. Πιο συγκεκριμένα, έστω ο στοχαστικός κατά γραμμές πίνακας (δηλαδή πίνακας για τον οποίο το άθροισμα κάθε γραμμής είναι μονάδα) \mathcal{H} διάστασης $n \times k$ ο οποίος παίρνει αποκλειστικά τις τιμές 0 και 1 (ενδείκτριος πίνακας). Με βάση τα προηγούμενα, θέτουμε $\mathcal{U} = \mathcal{H}\mathcal{H}^\top$ και το πρόβλημα ελαχιστοποίησης γίνεται αντίστοιχα:

$$\min_{\mathcal{H}} \left\| \tilde{\mathcal{M}} - \mathcal{H}\mathcal{H}^\top \right\|^2 \text{ ώστε ο } \mathcal{H} \text{ να είναι ενδείκτριος}$$

ο παραπάνω περιορισμός είναι δύσκολος στην επίλυση του. Εναλλακτικά, θεωρούμε $\mathcal{H}^\top \mathcal{H} = [h_{i,j}]$ οπότε και εκφράζουμε τον περιορισμό ως $h_{i,j} = 0$ αν $i \neq j$ και $h_{i,i} = \eta_i$ (η χαλάρωση του περιορισμού οφείλεται στο ότι επιτρέπουμε $\eta_i \neq 1$). Αν ορίσει κανείς

$$\mathcal{D} = \mathcal{H}^\top \mathcal{H} = \begin{pmatrix} \eta_1 & 0 & \cdots & 0 \\ 0 & \eta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \eta_k \end{pmatrix}$$

τότε η ελαχιστοποίηση μετατρέπεται στην

$$\mathcal{H}^\top \mathcal{H} = \mathcal{D}, \mathcal{H} \geq 0 \quad \left\| \tilde{\mathcal{M}} - \mathcal{H}\mathcal{H}^\top \right\|^2$$

η οποία θεωρεί το μέγεθος των cluster γνωστό, μέγεθος το οποίο δε μας είναι γνωστό προτού λύσει κανείς το ίδιο το πρόβλημα. Επομένως, χρειάζεται να απαλειφει κανείς τον πίνακα \mathcal{D} . Συνεπώς, ορίζει κανείς $\tilde{\mathcal{H}} = \mathcal{H}(\mathcal{H}^\top \mathcal{H})^{-\frac{1}{2}}$ οπότε:

$$\mathcal{H}\mathcal{H}^\top = \tilde{\mathcal{H}}\mathcal{D}\tilde{\mathcal{H}}^\top \text{ και } \tilde{\mathcal{H}}^\top \tilde{\mathcal{H}} = \mathcal{H}(\mathcal{H}^\top \mathcal{H})^{-1}\mathcal{H} = I$$

οπότε η ελαχιστοποίηση διαμορφώνεται τελικά ως

$$\tilde{\mathcal{H}}^\top \tilde{\mathcal{H}} = I, \tilde{\mathcal{H}}, \mathcal{D} \geq 0 \quad \left\| \tilde{\mathcal{M}} - \tilde{\mathcal{H}}\mathcal{D}\tilde{\mathcal{H}}^\top \right\|^2, \text{ με } \mathcal{D} \text{ διαγώνιο πίνακα}$$

Η απαίτηση ότι ο πίνακας \mathcal{D} είναι διαγώνιος συνήθως μετατρέπεται σε μια πιο χαλαρή απαίτηση του να είναι συμμετρικός και μη αρνητικός πίνακας. Αναλυτικές λεπτομέρειες υπάρχουν στα [31], [39], όπου προτάθηκε η παραπάνω ανάλυση.

4.5.3 Το scaling πρόβλημα και η εισαγωγή smoothness constraints - Μέρος 1

Όπως αναφέρεται και στο [31], η παραπάνω ιδέα είναι υπολογιστικά αδύνατη για εφαρμογές όπως η κατάτμηση εικόνας, καθώς για μια εικόνα $p \times q$ pixel απαιτούνται πολύ μεγάλοι πίνακες \mathcal{M}_t σε κάθε διαμέριση t . Επομένως, προτάθηκε στο [31] η συνένωση pixel τα οποία ανήκουν σε κάθε διαμέριση t στο ίδιο cluster και ταυτόχρονα είναι γειτονικά. Με τον τρόπο αυτό δημιουργούνται πλέον αντικείμενα (objects) τα οποία αποτελούν σύνολα pixel. Εφαρμόζοντας το Consensus Clustering αποδίδουμε labels όχι στα αρχικά pixel, αλλά στα αντικείμενα. Έπειτα, το label ενός αντικειμένου κληρονομείται σε όλα τα pixel αυτού. Με τον τρόπο αυτό, μειώνεται σημαντικά η διάσταση των πινάκων οι οποίοι υπεισέρχονται στην προηγούμενη ανάλυση από $p \times q$ σε $N_{obj} \times N_{obj}$ με N_{obj} τον αριθμό των αντικειμένων.

Επίσης, είναι δυνατή η ενσωμάτωση περαιτέρω constraints στην προηγούμενη ελαχιστοποίηση ως εξής:

$$\tilde{\mathcal{H}}^\top \tilde{\mathcal{H}} = I, \tilde{\mathcal{H}}, \mathcal{D} \geq 0 \quad \left\| \tilde{\mathcal{M}} - \tilde{\mathcal{H}}\mathcal{D}\tilde{\mathcal{H}}^\top \right\|^2 + \lambda_{con} Tr(\tilde{\mathcal{H}}^\top \Theta \tilde{\mathcal{H}}) \quad (4.22)$$

όπου ο Θ είναι διάστασης $N_{obj} \times N_{obj}$ και αναπαριστά τους επιπρόσθετους περιορισμούς, ενώ η παράμετρος λ_{con} αποτυπώνει το trade - off μεταξύ των 2 όρων της (4.22). Ορίζοντας $d_{i,j}$ την επικάλυψη των ορίων (σε αριθμό pixel) μεταξύ 2 αντικειμένων i και j , μπορούμε να υπολογίσουμε το $\Theta_{i,j}$ ως εξής:

$$\Theta_{i,j} = \frac{1}{1 + \exp\left(-\frac{1}{d_{i,j} + \epsilon}\right)} \quad \text{όπου} \quad \begin{cases} \epsilon = 0.1 & \text{αν } \nexists \text{ επικάλυψη} \\ \epsilon = 0 & \text{αν } \exists \text{ επικάλυψη} \end{cases}$$

4.5.4 Μαθηματική επίλυση του Consensus Learning

Με βάση τα παραπάνω και την ανάλυση στα [31] και [36], οι εξισώσεις για την ελαχιστοποίηση της (4.22) είναι ($\tilde{\mathcal{H}} = [\tilde{h}_{i,j}]$ με αρχικοποίηση τη λύση του K-Means στον $\tilde{\mathcal{M}}$):

$$\mathcal{D} = (\tilde{\mathcal{H}}^\top \tilde{\mathcal{H}})^{-1} \tilde{\mathcal{H}}^\top \tilde{\mathcal{M}} \tilde{\mathcal{H}} (\tilde{\mathcal{H}}^\top \tilde{\mathcal{H}})^{-1} \quad (4.23)$$

$$\tilde{h}_{i,j} = \tilde{h}_{i,j} \sqrt{\frac{(\tilde{\mathcal{M}}\tilde{\mathcal{H}}\mathcal{D})_{ij}^+ + (\tilde{\mathcal{H}}(\mathcal{D}\tilde{\mathcal{H}}^\top\tilde{\mathcal{H}}\mathcal{D})^-)_{ij}}{(\tilde{\mathcal{M}}\tilde{\mathcal{H}}\mathcal{D})_{ij}^- + (\tilde{\mathcal{H}}(\mathcal{D}\tilde{\mathcal{H}}^\top\tilde{\mathcal{H}}\mathcal{D})^+)_{ij} + (\Theta\tilde{\mathcal{H}})_{ij}}} \quad (4.24)$$

$$\tilde{\mathcal{M}}\tilde{\mathcal{H}}\mathcal{D} = (\tilde{\mathcal{M}}\tilde{\mathcal{H}}\mathcal{D})^+ - (\tilde{\mathcal{M}}\tilde{\mathcal{H}}\mathcal{D})^- \quad (4.25)$$

$$\mathcal{D}\tilde{\mathcal{H}}^\top\tilde{\mathcal{H}}\mathcal{D} = (\mathcal{D}\tilde{\mathcal{H}}^\top\tilde{\mathcal{H}}\mathcal{D})^+ - (\mathcal{D}\tilde{\mathcal{H}}^\top\tilde{\mathcal{H}}\mathcal{D})^- \quad (4.26)$$

όπου ο υπερδείκτης + ή - δηλώνει το θετικό (και αντίστοιχα το αρνητικό κομμάτι) του πίνακα. Με τη βοήθεια των (4.25), (4.26) επαναλαμβάνει κανείς τις εξισώσεις (4.23), (4.24) εναλλάξ για έναν προκαθορισμένο αριθμό επαναλήψεων.

4.5.5 Το scaling πρόβλημα και η εισαγωγή smoothness constraints - Μέρος 2

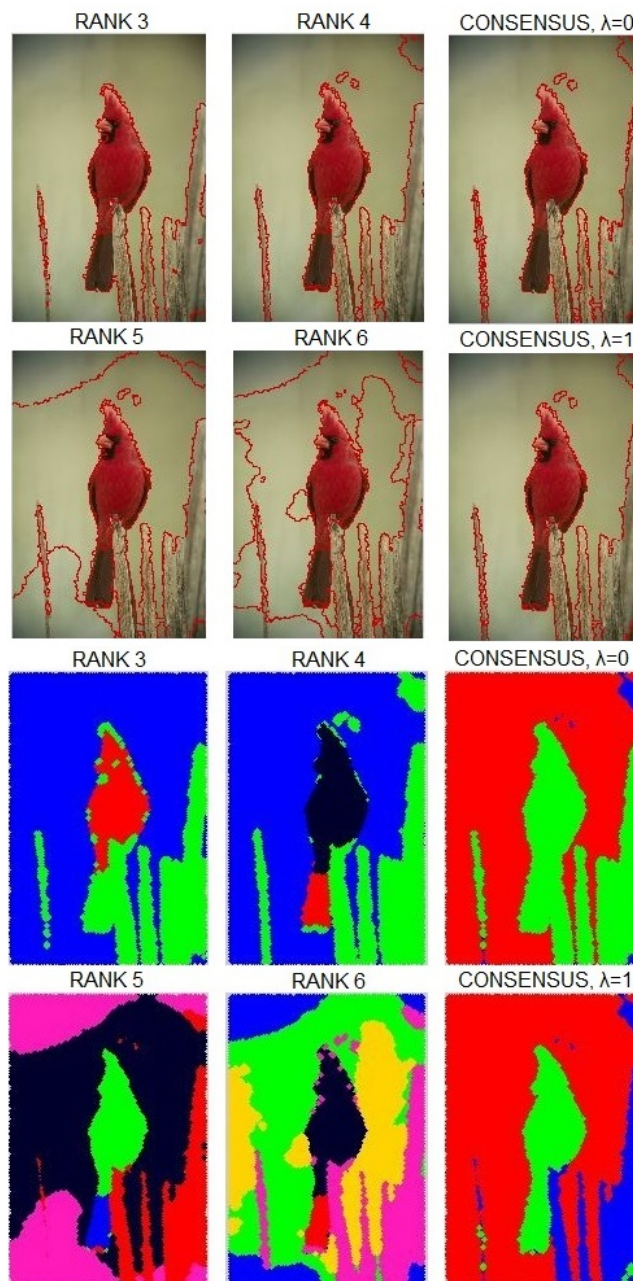
Μια ενδιαφέρουσα επέκταση του παραπάνω είναι να θεωρήσει κανείς την περίπτωση όπου η εικόνα αναπαρίσταται όχι με regular grid αλλά με Region Adjacency γράφο. Στην περίπτωση αυτή, δεν μιλάμε για συνένωση γειτονικών pixel σε αντικείμενα, αλλά για συνένωση κόμβων σε αντικείμενα. Προφανώς, όταν ένας κόμβος εκφυλίζεται σε pixel ισχύουν όσα ανέφερα προηγουμένως. Στη γενικότερη περίπτωση, καταφέρνουμε να εισάγουμε ένα ακόμη αφαιρετικό επίπεδο στην παραπάνω ανάλυση, με αποτέλεσμα να μειώνουμε ακόμα πιο δραστικά τον αριθμό των αντικειμένων. Έτσι, μπορούμε να επεξεργαστούμε εικόνες ακόμα μεγαλύτερου μεγέθους (πχ 640 × 640) αλλά και τις πιο μικρές ακόμα πιο γρήγορα.

Κατ' αντιστοιχία με τους παραπάνω smoothness constraints, μπορούμε να ορίσουμε σαν $d_{i,j}$ τον αριθμό των εξερχόμενων ακμών από ένα αντικείμενο i (σύνολο κόμβων) προς ένα άλλο αντικείμενο j . Ο αριθμός αυτός προκύπτει από το σύνολο των ακμών των κόμβων των 2 αντικειμένων κάθε φορά.

4.5.6 Color/Texture Consensus Learning

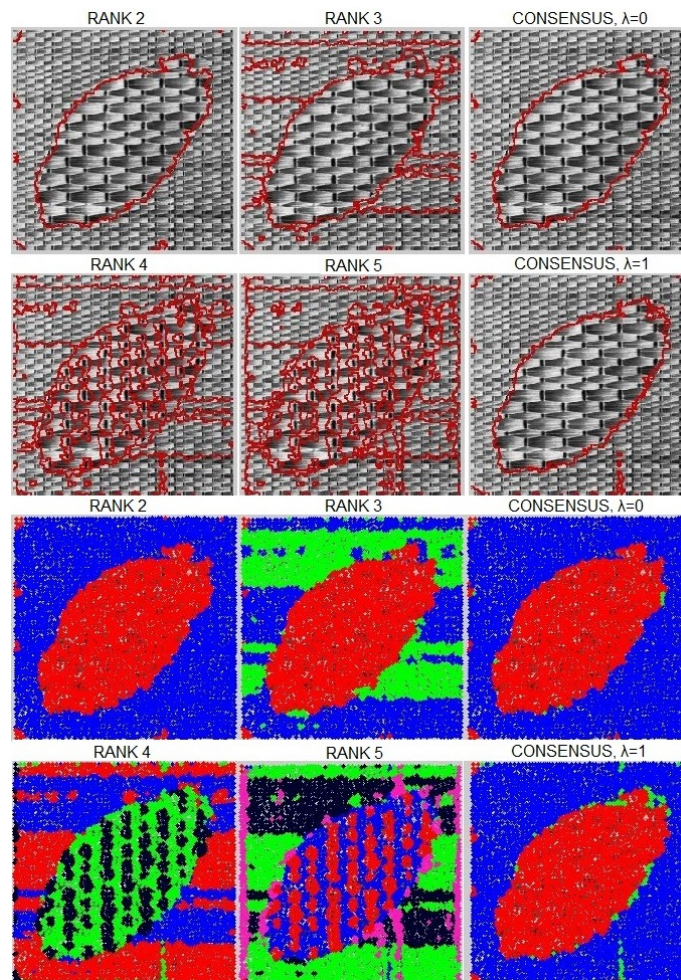
Με βάση τα παραπάνω, μπορούμε να συνδυάσουμε επιμέρους διαμερίσεις μιας εικόνας (σε διαφορετικό αριθμό parts) και να αποδώσουμε μια συνολική κατάτμηση της εικόνας. Ένα ακόμη πλεονέκτημα είναι ότι εξασφαλίζουμε ότι ακόμα και όταν κάποια εκ των αρχικών κατατμήσεων να μην είναι καλή (είτε λόγω χαμηλής ποιότητας των χαρακτηριστικών είτε λόγω του ίδιου του Cluster NMF το οποίο ενδεχομένως να δώσει μεταβαλλόμενα αποτελέσματα λόγω της αρχικοποίησης), το τελικό αποτέλεσμα να τη σταθμίζει και με βάση τις υπόλοιπες. Τέλος, μειώνεται η επίδραση της αρχικής επιλογής του βαθμού της λύσης, αφού κανείς συνδυάζει πολλαπλούς βαθμούς λύσεων με τον τρόπο αυτό σε μια ενιαία λύση.

Επιπρόσθετα, μπορεί κανείς αντί να αποκτή αποτελέσματα μόνο σε επίπεδο κόμβων, να τα μετατρέπει σε επίπεδο pixel. Προς τούτο, αποδίδουμε σε όλα τα pixel ενός κόμβου την ετικέτα του κόμβου που τους αντιστοιχεί. Για όλα τα pixel τα οποία αποτελούν region boundaries, επιλέγουμε σαν ετικέτα την ετικέτα της περιοχής στην οποία ανήκει η πλειοψηφία των 8 γειτόνων καθενός pixel. Με τη χρήση πινάκων και την αποδοτική υλοποίηση, η παραπάνω διαδικασία εισάγει έναν μη αμελητέο αλλά όχι ιδιαίτερα σημαντικό χρόνο εκτέλεσης. Για την περίπτωση της κατάτμησης με την πληροφορία του χρώματος αποκλειστικά, χρησιμοποιούμε το χρωματικό χώρο Lab σε διαφορετικό βαθμό λύσης κάθε φορά.



Σχήμα 4.31: Color Segmentation (pixel + node level), αρχικά rank: 3, 4, 5, 6 τελικό rank: 3, $\lambda_{base} = 0.02$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, bird.

Από το παραπάνω σχήμα, βλέπουμε ότι η εισαγωγή του χωρικού όρου όχι μόνο στις επιμέρους κατατμήσεις αλλά και στην τελική, βοηθάει στο να απομονώσει κανείς επιτυχώς το πουλί από τα ξύλα τα οποία βρίσκονται στο background. Ακόμα, μπορεί κανείς να χρησιμοποιήσει το σύστημα αυτό και στην κατάτμηση υφής. Η μόνη διαφορά με πριν, είναι ο χώρος των χαρακτηριστικών όπου και χρησιμοποιούμε τα χαρακτηριστικά Gabor, όπως έγινε και σε προηγούμενη ενότητα.



Σχήμα 4.32: Texture Segmentation (pixel + node level), αρχικά rank: 2, 3, 4, 5, τελικό rank: 3, $\lambda_{base} = 0.02$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, Brodatz 4.

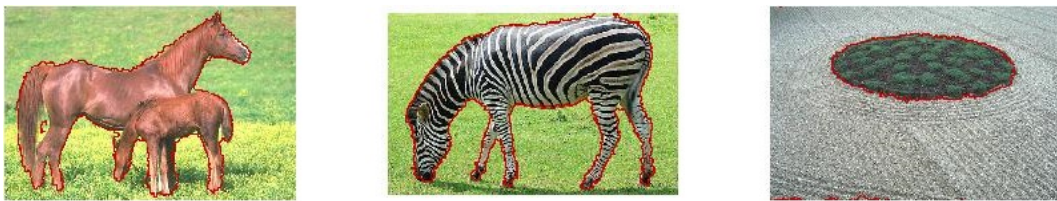
4.5.7 Συνολικό σύστημα

Συνδυάζοντας τις παραπάνω ιδέες, εισάγουμε το τελικό σύστημα κατάτμησης εικόνων ως εξής: επιτρέπουμε την κοινή χρήση χαρακτηριστικών χρώματος και υφής σε διάφορους βαθμούς αναπαράστασης (ranks). Με τον τρόπο αυτό ελαττώνεται η βαρύτητα στα χαρακτηριστικά και στο βαθμό αναπαράστασης, ενώ τονίζεται η προσπάθεια του να πετύχει κανείς μια κατάτμηση η οποία να συνδυάζει όλους τους βαθμούς αναπαράστασης και για τις 2 κατηγορίες χαρακτηριστικών (χρώματος και υφής).

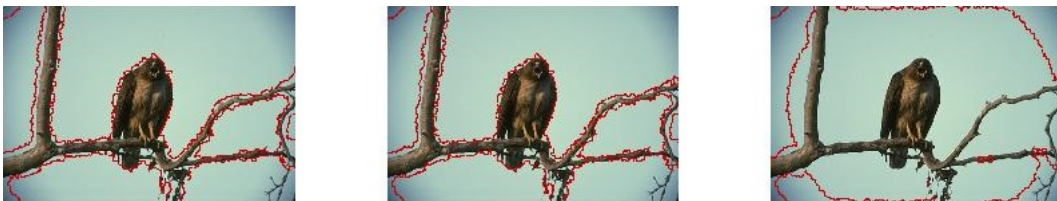
Επίσης, χρησιμοποιούμε χαρακτηριστικά υφής προσαρμοσμένα στην κατάτμηση φυσικών

εικόνων (natural images). Για το λόγο αυτό, θεωρούμε τον τελεστή $\Psi(f)$ ο οποίος μοντελοποιεί την ενέργεια της αρχικής εικόνας φιλτραρισμένης κάθε φορά με το αντίστοιχο φίλτρο Gabor. Στην περίπτωση αυτή, επιλέγουμε 10 προσανατολισμούς και 4 κλίμακες ώστε να προκύψουν 40 εκδοχές φιλτραρισμένων εικόνων. Ο τελεστής $\Psi(f)$ ορίζεται ως: $\Psi(f) = \|\nabla f\|^2 - f\nabla^2 f$.

Με τη βοήθεια του τελεστή $\Psi(f_{ij})$ όπου $f = I * g_{ij} \forall i = 1 \dots 4, j = 1 \dots 10$, υπολογίζουμε τα 5 κύρια φίλτρα τα οποία συνεισφέρουν κατά κύριο λόγο, δίνουν, δηλαδή, τη μεγαλύτερη ενέργεια. Η συνεισφορά αυτής της επιλογής είναι διπλή: μειώνουμε το χώρο χαρακτηριστικών της υψής από $40 \times n$ σε $5 \times n$ και εξασφαλίζουμε ότι χρησιμοποιούμε έναν αριθμό αποχρίσεων οι οποίες όντως συνεισφέρουν στο τελικό αποτέλεσμα.



Σχήμα 4.33: Συνολικό σύστημα (pixel level, αρχικά rank: 3(Y), 4(Y), 3(X), 4(X), τελικό rank: 2, $\lambda_{con} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = 0$, horses, zebra, grass.



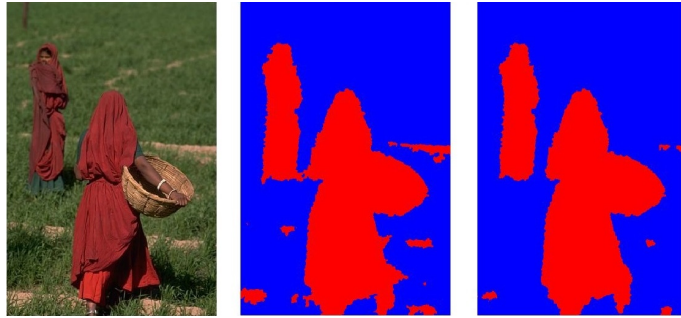
Σχήμα 4.34: Συνολικό σύστημα και μεταβλητότητα αποτελέσματος (pixel level, αρχικά rank: 3(Y), 4(Y), 3(X), 4(X), τελικό rank: 2, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = 0$, eagle.



Σχήμα 4.35: Συνολικό σύστημα και μεταβλητότητα αποτελέσματος (pixel level, αρχικά rank: 3(Y), 4(Y), 3(X), 4(X), τελικό rank: [2,3,4], $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = 1$, kangaroo.



Σχήμα 4.36: Δύσκολες περιπτώσεις (pixel level, αρχικά rank: 3(Υ), 4(Υ), 3(X), 4(X), τελικό rank: 2, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = 1$, kangaroo, tiger, small bear.



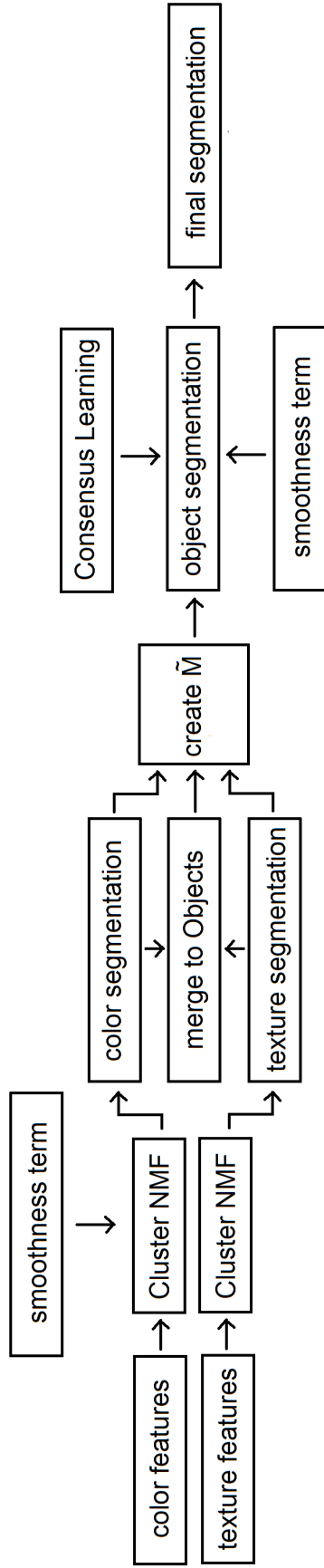
Σχήμα 4.37: Βελτίωση λόγω χωρικού όρου λ_{con} σε χρώμα μόνο (pixel level, αρχικά rank: 3(X), 4(X), 3(X), 4(X), τελικό rank: 2, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, women.



Σχήμα 4.38: Βελτίωση λόγω χωρικού όρου λ_{con} (pixel level, αρχικά rank: 3(Υ), 4(Υ), 3(X), 4(X), τελικό rank: 3, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, small bird.



Σχήμα 4.39: Πρόσθετο αποτέλεσμα (pixel level, αρχικά rank: 3(Υ), 4(Υ), 3(X), 4(X), τελικό rank: 5, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, woman + hair.



Σχήμα 4.40: Σχεδιάγραμμα συνολικού συστήματος.

4.6 Συμπεράσματα κεφαλαίου 4 - Ανακεφαλαίωση

Στο κεφάλαιο αυτό, έγινε έντονη ανάλυση της μεθόδου Cluster (NMF) με σκοπό την ανίχνευση αντικειμένων. Ξεκινώντας την ανάλυση από τον διαδεδομένο αλγόριθμο K-Means και την τεχνική NMF, παρατηρήσαμε μια έντονη εξάρτηση από την αρχικοποίηση και τον αριθμό των κλάσεων που κανείς ζητάει σαν τελικό αποτέλεσμα της μεθόδου. Η παραλλαγή του Cluster (NMF) με αρχικοποίηση μια λύση του K-Means έδειξε ότι μπορεί να ελαττώσει την αρνητική αυτή συγκυρία.

Όμως, πέραν αυτού, παρατηρήθηκε και πειραματικά ότι η χρήση κάποιων χαρακτηριστικών δίχως να εισάγεται μια γεωμετρική διάσταση στη μέθοδο, έδινε λύσεις με αρκετά θορυβώδη σημεία τα οποία διασκορπίζονται στην εικόνα. Για το λόγο αυτό, χρησιμοποιήθηκαν τεχνικές εισαγωγής χωρικών όρων ώστε γειτονικά συνδεδεμένοι κόμβοι να ενθαρρύνονται στην από κοινού απόκτηση μιας ετικέτας. Με τον τρόπο αυτό, επετεύχθη η κατάτμηση εικόνων οι οποίες είχαν ορίσει πάνω σε RAG είτε με τη χρήση χρωματικής πληροφορίας είτε με τη χρήση πληροφορίας υφής.

Οι πραγματικές εικόνες, όμως, χαρακτηρίζονται από ένα συνδυασμό πληροφορίας χρώματος και υφής. Επομένως, εξερευνήθηκαν ιδέες από Consensus Learning, όπου και συνδυάζει κανείς τις επιμέρους κατατμήσεις σε μια ενιαία. Τελικά, καταφέραμε να επιτύχουμε κατάτμηση φυσικών εικόνων με αρκετά καλά αποτελέσματα, τα οποία επηρεάζονται λιγότερο από την αρχικοποίηση (αφού υπάρχουν πολλές διαφορετικές κατατμήσεις διαθέσιμες) και από τον αριθμό των κλάσεων. Η βελτίωση των αποτελεσμάτων θα στηριχθεί κυρίως στη βελτίωση της διαδικασίας του feature extraction. Άλλωστε, κύριος σκοπός ήταν η ανάδειξη των προοπτικών της μεθόδου για κατάτμηση εικόνων και όχι τόσο ο προσδιορισμός των καλύτερων χαρακτηριστικών που θα το επιτύχει.

Κεφάλαιο 5

Σύγκριση μεταξύ Unsupervised και Supervised Τεχνικών

Στο κεφάλαιο αυτό, θα αναφερθούμε στις διαφορές μεταξύ των Supervised (Random Walker) και Unsupervised (NMF) τεχνικών τις οποίες είδαμε αντίστοιχα στα κεφάλαια 3 και 4.

5.1 Τεχνικές με επίβλεψη (Supervised)

5.1.1 Πλεονεκτήματα Supervised τεχνικών

1. Σε πολλές εφαρμογές είναι επιθυμητή η ενσωμάτωση πληροφορίας από τον χρήστη με σκοπό τη διαδραστική επεξεργασία των αποτελεσμάτων. Συνηθίζεται, επομένως, η απόδοση κάποιων αρχικών seeds από το χρήστη και παρουσίαση σε αυτόν των αντίστοιχων αποτελεσμάτων. Έπειτα, αν η ποιότητα του αποτελέσματος δεν κρίνεται ικανοποιητική, τότε ο χρήστης μπορεί να δώσει σαν είσοδο περισσότερους σπόρους ή σε διαφορετική θέση ωστόσο να πάρει το ζητούμενο αποτέλεσμα.
2. Με δεδομένη την πληροφορία που μας έχει δώσει ο χρήστης, τα αποτελέσματα δεν αλλάζουν από εκτέλεση σε εκτέλεση. Με άλλα λόγια, δεν υπάρχει εξάρτηση από αρχικές συνθήκες ή κάποιο σύνολο λύσεων στο οποίο το τελικό αποτέλεσμα θα βρίσκεται. Η τελική κατάτμηση είναι μοναδική.

5.1.2 Μειονεκτήματα Supervised τεχνικών

1. Τα δεδομένα από τους χρήστες είναι συνήθως μικρής ποσότητας και δυσεύρετα. Επίσης, η αξιοπιστία αυτών συνδέεται αρκετά με το τελικό αποτέλεσμα, οπότε και μπορεί σε δυσμενή περιβάλλοντα να υπάρχουν έντονες αποκλίσεις λόγω κακής ποιότητας προσημασμένων δεδομένων.
2. Ο αριθμός και η θέση των αρχικών σπόρων επιδρά καίρια στο τελικό αποτέλεσμα, δίνοντας διαφορετικά αποτελέσματα ανάλογα με την επιλογή αυτών. Επίσης, κάθε χρήστης μπορεί να δίνει διαφορετικά σημεία σπόρους διότι μπορεί να θεωρήσει σημαντικά σημεία σε μια περιοχή διαφορετικά από ότι ένας άλλος χρήστης.

3. Απαιτείται πρότερη γνώση για τη δημιουργία των w_{ij} ή ισοδύναμα για τον προσδιορισμό του εύρους σ . Αν αυτή δεν είναι υπαρκτή, οι μέθοδοι οι οποίες την εξασφαλίζουν προσθέτουν χρόνο υπολογιστικό και πολυπλοκότητα στο τελικό σύστημα.
4. Η παράμετρος σ όχι μόνο χρήζει προσδιορισμού αλλά επιδρά σημαντικά και στο τελικό αποτέλεσμα. Η ευαισθησία αυτή αντικατοπτρίζεται έντονα και στον αλγόριθμο Random Walker και στο label propagation.

5.2 Τεχνικές δίχως επίβλεψη (Unsupervised)

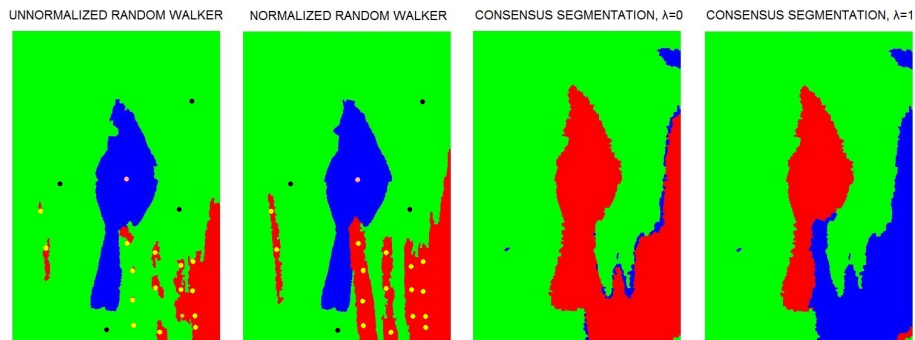
5.2.1 Πλεονεκτήματα Unsupervised τεχνικών

1. Δε χρειαζόμαστε σημεία σπόρους (seeds) παρά μόνο τον αριθμό των clusters (όπως ακριβώς δηλαδή και στον K-Means) στα οποία θέλουμε να χωρίσουμε τα δεδομένα. Έτσι, δεν υπάρχει επίδραση της θέσης/αριθμού αυτών.
2. Δε χρειάζεται η επίλυση γραμμικού (-ων) συστήματος (-των) όπως στον αλγόριθμο Random Walker ούτε η αντιστροφή του πίνακα L για label propagation τεχνικές. Η παραπάνω διαδικασία είναι ιδιαίτερα χρονοβόρα όταν χρησιμοποιεί κανείς 4/8 regular grid και μας περιορίζει σε σχετικά μικρές εικόνες.
3. Δεν απαιτείται καμία παράμετρος για τη δημιουργία των w_{ij} . Οι συσχετισμοί μεταξύ των κόμβων προσδιορίζονται κατά την ελαχιστοποίηση των αντίστοιχων συναρτησιακών κάθε μεθόδου.

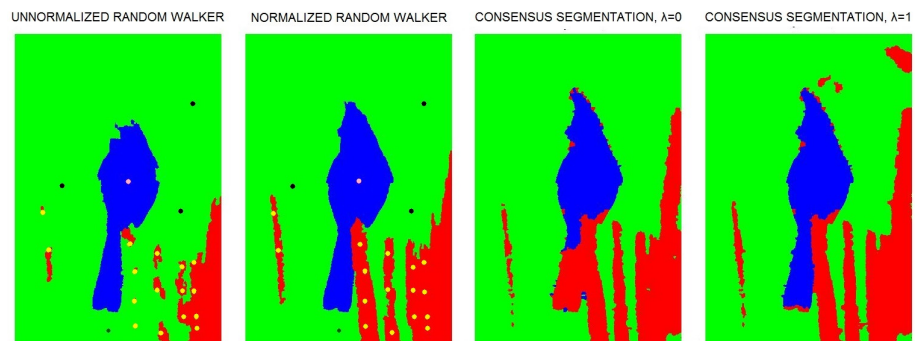
5.2.2 Μειονεκτήματα Unsupervised τεχνικών

1. Οι αλγόριθμοι NMF συνήθως έχουν τυχαίες αρχικοποιήσεις με αποτέλεσμα να είναι ευαίσθητοι στην αρχικοποίηση και να απαιτούνται αρκετές επαναλήψεις για τη σύγκλιση ([19]). Ακόμα και να μειώσει κανείς τον πιθανό χώρο λύσεων (όπως κάναμε με τη χρήση μιας K-Means αρχικοποίησης), χρειάζεται πάντα να λαμβάνεται υπόψη η ευαισθησία αυτή κυρίως όταν ο χώρος χαρακτηριστικών δεν είναι και ο πλέον κατάλληλος.
2. Η απόδοση του τελικού αποτελέσματος εξαρτάται από το βαθμό της αναπαράστασης που επιλέγουμε (ακόμα και στην περίπτωση του Consensus Learning το τελικό αποτέλεσμα δίνεται με βάση ένα συγκεκριμένο αριθμό αναπαραστάσεων - parts).
3. Ο μέσος χρόνος εκτέλεσης είναι μεγαλύτερος συγκριτικά με τις Supervised τεχνικές. Ο χρόνος αυτός αντιστοιχεί στην αντιστάθμιση της επιπρόσθετης πληροφορίας που δίνει ο χρήστης (όπως τα σημεία σπόροι ή η παράμετρος σ).
4. Το γεγονός ότι δεν απαιτείται κάποια παράμετρος (όπως το εύρος σ) έχει και αρνητικές επιπτώσεις. Επειδή ακριβώς οι τεχνικές NMF συνδέονται με τον αλγόριθμο K-Means, ο οποίος στερείται γεωμετρικής ενσωμάτωσης, απαιτήθηκε να εισάγουμε χωρικούς όρους. Οι όροι αυτοί ($\lambda_{spatial}$ ή λ_{base} για τις επιμέρους καταταμήσεις και λ_{con} στην τελική κατάταξη του Consensus Learning) υπονοούν το trade-off μεταξύ της

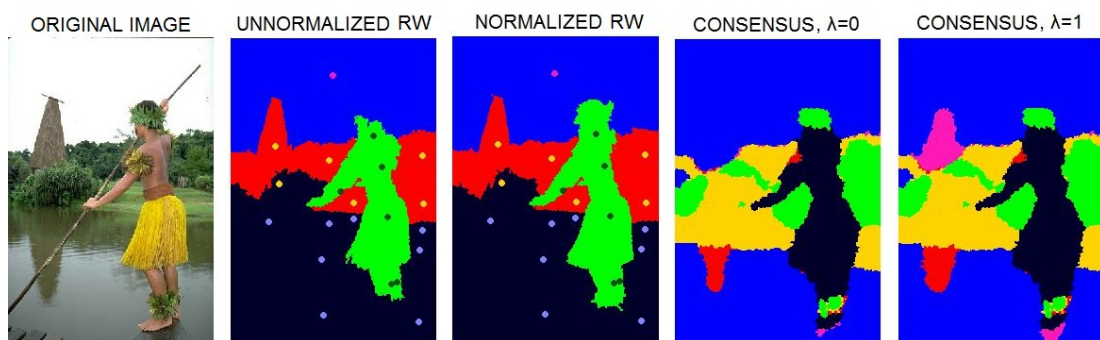
λύσης δίχως χωρικούς όρους και της επίδρασης αυτών. Συμπερασματικά, εισάγονται κάποιες παραπάνω παράμετροι οι οποίες επιδρούν και αυτές με τη σειρά τους στο τελικό αποτέλεσμα.



Σχήμα 5.1: Σύγκριση (pixel level), $\sigma_{RW} = 90$, αρχικά rank: 3(X), 4(X), 5(X), 6(X), τελικό rank: 3, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, bird.



Σχήμα 5.2: Σύγκριση (pixel level), $\sigma_{RW} = 100$, αρχικά rank: 3(X), 4(X), 5(X), 6(X), τελικό rank: 3, $\lambda_{base} = 0.02$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, bird.



Σχήμα 5.3: Σύγκριση (pixel level), $\sigma_{RW} = 90$, αρχικά rank: 5(X), 6(X), 5(X), 6(X), τελικό rank: 6, $\lambda_{base} = 0.2$, 150 επαναλήψεις όλα τα στάδια, $\lambda_{con} = [0, 1]$, boatman.

Κεφάλαιο 6

Συμβολές της Διπλωματικής Εργασίας και Μελλοντική Έρευνα

6.1 Συμβολές της διπλωματικής εργασίας

Η συγκεκριμένη διπλωματική εργασία εστίασε στη χρήση γράφων μη κανονικής δομής με σκοπό την ανίχνευση αντικειμένων. Για το σκοπό αυτό, μελετήθηκαν δύο κυρίαρχα είδη τεχνικών, τεχνικές με επίβλεψη και τεχνικές δίχως επίβλεψη. Οι κυριότερες συμβολές της εργασίας ήταν:

Στο κομμάτι των τεχνικών με επίβλεψη, έγινε θεωρητική σύνδεση του επιδημιολογικού μοντέλου SIR με τη διάχυση της θερμότητας πάνω σε γράφους. Η σύνδεση αυτή μας επέτρεψε τη μελέτη του label propagation με τη χρήση των αντίστοιχων μερικών διαφορικών εξισώσεων (με προκαθορισμένο αριθμό επαναλήψεων αλλά και στη μόνιμη κατάσταση). Αντιμετωπίζοντας τη μόνιμη κατάσταση με την ισοδύναμη μορφή της (Random Walker αλγόριθμος), υλοποιήθηκαν οι αλγόριθμοι Random Walker, Power Watershed σε επίπεδο κόμβων με τη χρήση RAG και αντίστοιχα προτάθηκε και υλοποιήθηκε ο αλγόριθμος Unbiased Random Walker. Στα πλαίσια της ελαχιστοποίησης συναρτησιακών, υλοποιήθηκε και το regularization σε έγχρωμες/grayscale εικόνες.

Συγκεντρώνοντας τις υλοποιήσεις αυτές, έγιναν πειραματικές συγκρίσεις μεταξύ των Random Walker και Unbiased Random Walker από άποψη χρόνου και ποιότητας τελικής κατάτμησης, είτε χρησιμοποιεί κανείς 4/8 regular grid είτε RAG. Η σύγκριση έγινε και σε επίπεδο ακρίβειας με βάση ένα συγκεκριμένο ground truth segmentation (στο 4/8 regular grid) αλλά και οπτικά (στο 4/8 regular grid και στο RAG). Επίσης, έγινε και μια γενική σύγκριση υπολογιστικού χρόνου μεταξύ 4/8 regular grid και RAG ανεξάρτητα από τον αλγόριθμο που χρησιμοποιεί κανείς.

Ένα βασικό κομμάτι για τις παραπάνω τεχνικές ήταν και ο προσδιορισμός των βαρών επί των ακμών μεταξύ των κόμβων. Στη βιβλιογραφία, έχουν μελετηθεί οι μέθοδοι LNP, GLNP προς αυτήν την κατεύθυνση. Υλοποιήθηκαν, επομένως, οι τεχνικές αυτές και επιχειρήθηκε ένας συνδυασμός του LNP με τον αλγόριθμο του Τυχαίου Περιπατητή. Πέραν αυτών, προσεγγίστηκε και ένας αυτόματος τρόπος κατάτμησης αντικειμένων πάνω σε γράφους στηριγμένος στη βασική συνάρτηση ομοιότητας (RBF kernel) που προτείνεται

συνήθως στις μεθόδους με επίβλεψη σε συνδυασμό με ιδέες από Network Science.

Στο κύριο κομμάτι των τεχνικών δίχως επίβλεψη, έγινε σύνδεση της μεθόδου GLNP με τη μέθοδο Cluster NMF καθώς και η υλοποίηση αυτών. Με τον τρόπο αυτό, προσεγγίσαμε εφαρμογές όπως η κατάτμηση υφής και η κατάτμηση φυσικών εικόνων μέσω του Cluster NMF. Παρατηρούμε ότι στη βιβλιογραφία, η κατάτμηση υφής με τη χρήση τεχνικών NMF έχει μελετηθεί σε περιορισμένη έκταση ([37]). Ακόμα, επιχειρήθηκε ο συνδυασμός λίγων regularization steps πριν την ελαχιστοποίηση με Cluster NMF. Στην πορεία της εργασίας, παρατηρήθηκε ότι οι λύσεις που ζητάει κανείς από τέτοιες μεθόδους χρειάζεται να ενσωματώνουν και τη δομή πάνω στο γράφο, δηλαδή και χωρικούς όρους.

Σαν αποτέλεσμα αυτής της ανάγκης, έγινε μια μαθηματική ανάλυση του Cluster NMF με σκοπό την ενσωμάτωση χωρικού όρου, η οποία καθοδήγηθηκε και από τη μελέτη του regularization στο προηγούμενο κεφάλαιο. Ο χωρικός αυτός όρος θέτει βάρη 0 ή 1 (όπως ένας πίνακας γειτνίασης) επί του RAG. Τέλος, στα πλαίσια της χρήσης παραπάνω από 1 ροής πληροφορίας, υλοποιήθηκε η μέθοδος Multi-View Learning ([20]) και το Consensus Learning ([31]) σε επίπεδο αντικειμένων ορισμένων επί κόμβων και όχι απλά επί γειτονικά συνδεδεμένων pixel.

Τέλος, σημειώνεται ότι σε διάφορα στάδια της εργασίας απαιτήθηκε υπολογιστικός κώδικας για την κατασκευή του RAG και τη μετατροπή των αποτελεσμάτων από το επίπεδο των κόμβων σε επίπεδο pixel (καθώς και για ακόμα ένα επίπεδο από αντικείμενα σε κόμβους στο Consensus Learning) για την εξαγωγή των περιγραμμάτων των αντικειμένων και των αντίστοιχων ετικετών για τα pixel. Για τις διεργασίες αυτές απαιτήθηκε μια αποτελεσματική υλοποίηση αυτών ώστε να επιτελούνται σε λογικά χρονικά πλαίσια.

6.2 Μελλοντική έρευνα

1. Ισχυροποίηση της υπεροχής του αλγόριθμου Unbiased Random Walker έναντι του παραδοσιακού αλγόριθμου. Παρά το γεγονός ότι στην πλειονότητα των περιπτώσεων προέκυψαν καλύτερα και πιο εύρωστα αποτελέσματα, μπορούμε να μελετήσουμε πιο διεξοδικά τη σύγκριση μεταξύ των δύο τεχνικών. Για παράδειγμα, είναι πιθανό ότι η εισαγωγή βελτιωμένων χαρακτηριστικών (και όχι απλά τα 3 κανάλια του χρωματικού χώρου RGB) να ενισχύσει την υπεροχή αυτή.
2. Ενσωμάτωση τεχνικών οι οποίες προσδιορίζουν την παράμετρο σ με αυτόματο τρόπο. Για παράδειγμα, μπορεί κανείς να υλοποιήσει τις μεθόδους στο [11], οι οποίες έχουν και τοπικό χαρακτήρα (δηλαδή δεν επιβάλλουν μια ολική σταθερά σαν τελική παράμετρο) αλλά και αναφέρονται σε κάθε χρωματικό κανάλι ξεχωριστά. Προφανώς, μπορούμε και να μελετήσουμε πιο διεξοδικά τη χρήση των LNP, GLNP προς την κατεύθυνση αυτή.
3. Ενσωμάτωση βελτιωμένων χαρακτηριστικών για το σχηματισμό των βαρών στον αλγόριθμο Unbiased Random Walker. Πέραν της βελτίωσης σε σχέση με τον παραδοσιακό αλγόριθμο (όπως προαναφέρθηκε), είναι εφικτή η επέκταση της μεθόδου για κατάτμηση υφής και φυσικών εικόνων με περισσότερο έντονο το στοιχείο της υφής, των οποίων η κατάτμηση είναι λιγότερο ακριβής αν χρησιμοποιεί κανείς μόνο τη

χρωματική πληροφορία.

4. Ενσωμάτωση βελτιωμένων χαρακτηριστικών για το σύστημα κατάτμησης φυσικών εικόνων, ώστε ακόμα και πιο δύσκολες εικόνες να μπορούν να κατατμηθούν επιτυχώς. Ενώ για τη χρωματική πληροφορία, επιλέχθηκε μια οικονομική και αρκετά καλή χρωματική αναπαράσταση (Lab), στο κομμάτι των χαρακτηριστικών υφης, υπάρχουν πολλές δυνατότητες με αποδεδειγμένα βελτιωτικά χαρακτηριστικά ([32]). Η βελτίωση του χώρου χαρακτηριστικών θα συμβάλλει θετικά στις επιμέρους κατατμήσεις, θα μειώσει τη μεταβλητότητα της μορφής τους και συνάμα θα βελτιώσει την τελική κατάτμηση. Τέλος, ο αριθμός των προκύπτοντων αντικειμένων στη διαδικασία του Consensus Learning θα μειωθεί περισσότερο, ελαττώνοντας και το συνολικό χρόνο εκτέλεσης.
5. Βελτίωση του χωρικού όρου για τις τεχνικές με το Cluster NMF. Η απόδοση binary βαρών αποδίδει με τον ίδιο τρόπο τη χωρική πληροφορία ανεξαρτήτως της ομοιότητας μεταξύ των 2 συνδεδεμένων κόμβων. Στο κομμάτι αυτό ιδέες από τον προσδιορισμό της παραμέτρου σ μπορούν να φανούν χρήσιμες.

Βιβλιογραφία

- [1] F. Wang, C. Zhang, "Label Propagation through Linear Neighborhoods", IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 1, pp. 55-67, Jan. 2008
- [2] S. Roweis, L. Saul, "Nonlinear dimensionality reduction by locally linear embedding", Science, vol. 290, no. 5500, pp. 2323-2326, Dec. 2000
- [3] Z. Tian, R. Kuang, "Global linear neighborhoods for efficient label propagation", Proc. of the SIAM Conference on Data Mining (SDM), SIAM/Omnipress, pp. 863-872, 2012
- [4] X. Zhu, Z. Ghahramani, J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions", Proc. of the 20-th International Conference on Machine Learning (ICML), 2003
- [5] X. Zhu, Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation", Technical Report 02-107, CMU-CALD, Jun. 2002
- [6] D. Lee, S. Seung, "Algorithms for non-negative matrix factorization", Proc. of NIPS, pp. 556-562, 2000
- [7] L. Grady, "Random walks for image segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 11, pp. 1768-1783, Nov. 2006
- [8] F. Nie, H. Wang, H. Huang, C. Ding, "Unsupervised and semi-supervised learning via ℓ_1 -norm graph", Proc. of the International Conference on Computer Vision (ICCV), pp. 2268-2273, Nov. 2011
- [9] Z. Lu, Y. Peng, "Robust Image Analysis by L1-Norm Semi-supervised Learning", CoRR, abs/1110.3109, Oct. 2011
- [10] A. Beck, M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems", SIAM Journal on Imaging Sciences, vol. 2, no. 1, pp. 183-202, 2009.
- [11] O. Lezoray, S. Bougleux, A. Elmoataz, "Parameterless discrete regularization on graphs for color image filtering", Proc. of the International Conference on Image Analysis and Recognition (ICIAR), 2007
- [12] E. Postnikov, I. Sokolov, "Continuum description of a contact infection spread in a SIR model", Mathematical Biosciences 208, pp. 205-215, 2007

- [13] D. Zhou, O. Bousquet, T. Lal, J. Weston, B. Schölkopf, "Learning with Local and Global Consistency", Proc. of NIPS, 2003
- [14] C. Couprie, L. Grady, L. Najman, H. Talbot, "Power Watershed: A Unifying Graph-Based Optimization Framework", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.33, no.7, pp. 1384-1399, July 2011
- [15] C. Ding, T. Li, M. Jordan, "Convex and Semi-Nonnegative Matrix Factorizations", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 1, pp. 45-55, Jan. 2010
- [16] Z. Cui, W. Li, G. Pan, L. Gao, "An Improved Random Walker using Spatial Feature for Image Segmentation", Proc. of the 25th Chinese Control and Decision Conference (CCDC), Guiyang, pp.1479-1482, 2013
- [17] G. Li, L. Qingsheng, C. Jian, "A New Fast Random Walk Segmentation Algorithm", Proc. of the 2nd International Symposium on Intelligent Information Technology Application, Shanghai, pp. 693-697, 2008
- [18] D. Zhou, B. Scholkopf, "Learning from Labeled and Unlabeled Data Using Random Walks", Proc. of the 26th DAGM Symposium, 2004
- [19] C. Boutsidis, E. Gallopoulos, "SVD based initialization: A head start for nonnegative matrix factorization", Technical Report HPCLAB-SCG-02/01-07, University of Patras, Greece
- [20] Z. Akata, C. Thurau, C. Bauckhage, "Non-negative Matrix Factorization in Multimodality Data for Segmentation and Label Prediction", Proc. of the 16th Computer Vision Winter Workshop, pp. 18, 2011.
- [21] D. Greene, P. Cunningham, "A matrix factorization approach for integrating multiple data views," Proc. of the Europ. Conf. Mach. Learn. Knowl. Discovery Datab. (ECML PKDD), Slovenia, pp. 423-438, 2009
- [22] W. Xu, X. Liu, Y. Gong, "Document clustering based on non-negative matrix factorization", Proc. of the 26th annual international ACM SIGIR conference, ACM Press, pp. 267-273, 2003
- [23] L. Grady, "Multilabel random walker image segmentation using prior models", Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 763-770, June 2005
- [24] Y. Artan, I. Yetik, "Improved Random Walker Algorithm for Image Segmentation", Proc. of the IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), pp. 89-92, May 2010
- [25] S. Andrews, G. Hamarneh, A. Saad, "Fast random walker with priors using pre-computation for interactive medical image segmentation", Proc. of Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 9-16, 2010

- [26] R. Courant, D. Hilbert, "Methods of Mathematical Physics", John Wiley and Sons, 1989, vol. 2.
- [27] P. Doyle, L. Snell, "Random walks and electric networks", ser. Carus mathematical monographs, Mathematical Association of America, no. 22, 1984
- [28] P. Bonacich, P. Lloyd, "Eigenvector-like measures of centrality for asymmetric relations", *Social Networks*, vol. 23, no. 3, pp. 191-201, 2001
- [29] D. Cai, X. He, J. Han, T. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, p. 1548-1560, Aug. 2011
- [30] A. Elmoataz, O. Lezoray, S. Bougleux, "Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing", *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1047-1060, Jul. 2008
- [31] S. Ghosh, J. Pfeiffer, J. Mulligan, "A general framework for reconciling multiple weak segmentations of an image", *Proc. of the IEEE Workshop on Applications of Computer Vision (WACV)*, Snowbird, pp. 1-9, Dec. 2009
- [32] I. Kokkinos, G. Evangelopoulos, P. Maragos, "Texture Analysis and Segmentation Using Modulation Features, Generative Models and Weighted Curve Evolution", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 142-157, Jan. 2009
- [33] V. Struc, N. Pavesic, "From Gabor Magnitude to Gabor Phase Features: Tackling the Problem of Face Recognition under Severe Illumination Changes", *Face Recognition*, Milos Oravec (Ed.), ISBN: 978-953-307-060-5, InTech, Apr. 2010
- [34] Y. Boykov, M.-P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images", *Proc. of the International Conference on Computer Vision (ICCV)*, pp. 105-112, 2001
- [35] J. Cousty, G. Bertrand, L. Najman, M. Couprie, "Watershed Cuts: Minimum Spanning Forests and the Drop of Water Principle", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 8, pp. 1362-1374, Aug. 2009
- [36] F. Wang, T. Li, C. Zhang, "Semi-supervised clustering via matrix factorization", *Proc. of SIAM Conference on Data Mining (SDM)*, pp. 1-12, 2008
- [37] J. Yuan, D. Wang, "Factorization-Based Texture Segmentation", Technical Report OSUT-CISRC-1/13-TR01, Department of Computer Science and Engineering-The Ohio State University, 2013
- [38] Y. Wang, Y. Jia, C. Hu, M. Turk, "Non-negative matrix factorization framework for face recognition", *Int. J. Pattern Recognition Artif. Intell.*, vol. 19, no. 4, pp. 495-511, 2005

- [39] T. Li, C.Ding, M. Jordan, "Solving Consensus and Semi-supervised Clustering Problems Using Nonnegative Matrix Factorization", Seventh IEEE International Conference on Data Mining (ICDM), pp. 577–582, Oct. 2007