



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΚΑΙ ΤΟΠΟΓΡΑΦΩΝ
ΜΗΧΑΝΙΚΩΝ

Δημιουργία Λογισμικού Ανοικτού Κώδικα για Κατάτμηση Ψευδέγχρωμων Δορυφορικών Δεδομένων σε Περιβάλλον Google Earth

Μεταπτυχιακή Διπλωματική του

Τεχνίτη Γεώργιου

Επιβλέπων Καθηγητής : Αργιαλάς Δημήτριος

Αθήνα, 2010

Περιεχόμενα

Περίληψη	6
Abstract	7
Ευχαριστίες	8
1 Εισαγωγή	9
2 Ανασκόπηση βιβλιογραφίας στην κατάτμηση εικόνας βασισμένη σε Mean-Shift- τμηματοποίηση δεδομένων	10
2.1 Εισαγωγή	10
2.2 Η διαδικασία Μέσης Μετατόπισης -Mean-Shift.....	13
2.2.1 Υπολογισμός της κλίσης της πυκνότητας	15
2.2.2 Επαρκής συνθήκη για σύγκλιση	17
2.2.3 Ανίχνευση κορυφών βασισμένη στη διαδικασία Mean Shift	19
2.2.4 Η ιδιότητα της ομαλής τροχιάς.....	22
2.2.5 Γενικές παραδοχές.....	22
2.2.6 Παράμετροι που θα πρέπει να ληφθούν υπόψη κατά την κατάτμηση.....	23
2.3 Δυναμική ανάλυση των χαρακτηριστικών χώρων	23
2.3.1 Επιλογή του εύρους.....	25
2.4 Εφαρμογές.....	26
2.4.1 Ομαλοποίηση που διατηρεί την ασυνέχεια (ακμές).....	28
2.4.2 Κατάτμηση εικόνας.....	32
3 Ελεύθερη κατανομή συσσώρευσης – θεωρία και εφαρμογή	35
3.1 Εισαγωγή	35
3.2 Ανάλυση Συσσώρευσης.....	36
3.3 Εκτίμηση απόδοσης.....	38
3.4 Κατάτμηση βασισμένη στα κέντρα συσσωρεύσεων.....	40
4 Google Earth Παρουσίαση περιβάλλοντος προγραμματισμού, εφαρμογή κατάτμησης στο Google Earth	45
4.1 Εισαγωγή	45
4.1.1 Η τεχνολογία του Google Earth	45
4.1.2 Παραμετροποιώντας το Google Earth.....	46

4.2	Το λογισμικό Earthwatch.....	46
4.2.1	Introduction	46
4.2.2	Τεχνικό Υπόβαθρο.....	48
4.2.3	Συνοπτικό Διάγραμμα Ροής Δεδομένων του EarthWatch	49
5	Μελέτες Περιπτώσεων Εφαρμογής του αλγορίθμου.....	54
5.1	Αεροδρόμιο Ελ. Βενιζέλος, Αθήνα.....	54
5.2	Big Island, Hawaii.....	59
5.3	Victoria Harbour, Hong Kong.....	62
5.4	Ο.Α.Κ.Α. Ολυμπιακές Εγκαταστάσεις, Αθήνα.....	69
5.5	Goa, Δυτική Ινδία.....	77
5.6	Εδιμβούργο, Ηνωμένο Βασίλειο	80
5.7	West Ghetto, Ναϊρόμπι.....	85
5.8	Πόλη της Βενετίας, Ιταλία	93
5.9	Ηράκλειο, Κρήτη	97
6	Αξιολόγηση - Συμπεράσματα – Μελλοντικές Εργασίες.....	101
6.1	Αξιολόγηση	101
6.1.1	Αεροδρόμιο Ελ. Βενιζέλος, Αθήνα.....	102
6.1.2	Big Island, Hawaii	103
6.1.3	Victoria Harbour, Hong Kong	104
6.1.4	Ο.Α.Κ.Α. Ολυμπιακές Εγκαταστάσεις, Αθήνα.....	105
6.1.5	Goa, Δυτική Ινδία.....	106
6.1.6	Εδιμβούργο, Ηνωμένο Βασίλειο.....	107
6.1.7	West Ghetto, Ναϊρόμπι	108
6.1.8	Πόλη της Βενετίας, Ιταλία.....	109
6.1.9	Ηράκλειο, Κρήτη	109
6.2	Συμπεράσματα για την κατάτμηση εικόνας	110
6.2.1	Γενικά Συμπεράσματα	110
6.2.2	Παρατηρήσεις.....	111
6.3	Συνολικά	111
6.4	Μελλοντικές Εργασίες.....	112

6.4.1	Βελτίωση του Αλγορίθμου Κατάτμησης.....	112
6.4.2	Βελτίωση – Ολοκλήρωση του συστήματος ανάλυσης	113
6.5	Επίλογος	114
7	Βιβλιογραφία	116
7.1	Διαδικτυακοί Χώροι.....	116
7.1.1	Γενικού Περιεχομένου	116
7.1.2	Αλγόριθμος Μέσης Μετατόπισης	116
7.1.3	Stitching images java source	117
7.1.4	GDal_wrapping_utility	118
7.1.5	Edison_Source_References.....	119
7.1.6	GE	119
8	Παράρτημα Ι.....	120
8.1	Earth Watch - Design and Implementation	120
8.1.1	Overall Design and Implementation	120
	2 OO.....	120
8.1.2	Data Model.....	120
8.1.3	Navigation	121
8.2	Learning Vector Quantization (LVQ).....	122
8.2.1	Learning Vector Quantization	122
8.2.2	Advantages of the Learning Vector Quantization algorithm	123
8.2.3	Disadvantages of the Learning Vector Quantization algorithm.....	123
8.2.4	Algorithm Notes	124
8.3	Programming Modules and technologies included in this project.....	127
8.3.1	Google Earth COM API	127
8.3.2	gekmlib.....	127
8.3.3	Stitching of the Images.....	127
8.3.4	Win32_COM reference	128
8.3.5	GDI 32 reference	130
8.3.6	Native Win32C using the Java Native Interface (JNI).....	131
9	Παράρτημα ΙΙ.....	133

9.1	Windows library reference - Shell32 (provides accessto system folder applications.	133
9.1.1	Shell32 library	133
9.1.2	User32 facilitates the simulation of the key strokes	134
9.2	Com4J - automatically generates the java classes that map directly to com interfaces ...	134
9.3	HWND - reference - provided by the com interface	134
9.4	win32GDI - to gain access to the display buffer	134

Περίληψη

Η διατριβή αυτή έχει ως κύριο στόχο να κάνει ευκολότερη την πρόσβαση, ενός χρήστη, σε ένα ανώτερο πακέτο ανάλυσης εικόνας, κυρίως για ερασιτεχνική επιστημονική μελέτη.

Η έλλειψη της διαλειτουργικότητας σήμερα, μειώνει το ρυθμό ανάπτυξης τεχνικών για άμεση ανάλυση εικόνας. Επί του παρόντος τα προγράμματα ανοικτού κώδικα εξαπλώνονται όλο και περισσότερο προς κάθε δυνατή κατεύθυνση, ενώ η δωρεάν διάδοση δεδομένων είναι επίσης μια βασική τάση στο σημερινό ρυθμό ζωής του γρήγορου διαδικτύου.

Η παρούσα διατριβή εισάγει έναν τρόπο να χρησιμοποιούνται τα δημοσιευμένα, και χωρίς χρέωση, δεδομένα στο διαδίκτυο, για τη διεξαγωγή πολυεπίπεδης κατάτμησης εικόνας. Αυτό γίνεται χρησιμοποιώντας τον αλγόριθμο μέσης μετατόπισης, σε συνδυασμό με ορισμένες βασικές ικανότητες κατηγοριοποίησης βασισμένες στον Learning Vector Quantization (WEKA) αλγόριθμο. Μακριά από το να γίνει εμπορικό προϊόν, αυτή η διατριβή προτείνει έναν απλό τρόπο για να αυξηθεί η συνέργεια τριών διαφορετικών πακέτων λογισμικού, προκειμένου να θέσει σε κίνηση περαιτέρω πιο οργανωμένες προσπάθειες προς αυτή την κατεύθυνση.

Η γνώση από μια σειρά αρκετά διαφορετικών τεχνολογιών, επιστρατεύτηκε για να επιτευχθεί το παρόν αποτέλεσμα. Από υπολογιστικά γραφικά, σε προχωρημένες στατιστικές μεθόδους για την ανάλυση εικόνας, μέχρι την ανάπτυξη περιβάλλοντος διάδρασης σε Java κτλ.

Η σύγκλιση της διαδικασίας μέσης μετατόπισης, προς την όσο το δυνατόν κοντινότερη απόδοση της υποκείμενης κατανομής έχει αποδειχθεί τόσο για τον πυρήνα Epanechnikov όσο και για το γενικής περίπτωσης πυρήνα με κυρτό και μονοτονικής μείωσης προφίλ. Η αξιοποίηση της διαδικασίας μέσης μετατόπισης εξελίσσει μια ισχυρή μέθοδο συσσώρευσης, που χρησιμοποιήθηκε για την ανάλυση των χώρων με πολύπλοκα χαρακτηριστικά που προέρχονται από RGB δορυφορικά δεδομένα. Τα κέντρα διασποράς που λαμβάνονται με την εξεύρεση της κατάστασης των υποκείμενων κατανομών και τις λεκάνες έλξης/επιρροής τους, καθορίζοντας έτσι τα όρια της συσσώρευσης. Γίνεται παρουσίαση παραδειγμάτων για την κατάτμηση της εικόνας, σε μια ποικιλία από εφαρμογές, αναδεικνύοντας έτσι, τις ανώτερες επιδόσεις του αλγορίθμου, καθώς και τον αβίαστο τρόπο να φιλοξενήσει επιστημονικά, ένα ευρύ φάσμα αναγκών στην ανάλυση εικόνας, από ένα απλό χρήστη που θέλει να διακρίνει το νερό από τη γη, ή να καταγράψει τα με πολύγωνα όλα τα κτίρια στο χωριό του, μέχρι και ένα έμπειρο χρήστη που προσπαθεί να εκτιμήσει την έκταση/επέκταση του οδικού δικτύου σε ένα αστικό χώρο, ή να υπολογισθεί η καμένη περιοχή μιας δασικής έκτασης.

Abstract

What this thesis aims to do is mainly to make a higher image-analysis package of information accessible for amateur scientific study.

The lack of the interoperability halts today the developing rate of the instant image analysis techniques. At present, open source programming is increasingly spreading towards any possible direction and the free data transfer is also a basic trend in the present fast internet- daily-lifestyle.

This thesis introduces a way to use the online free satellite data, in order to conduct multilevel segmentation analysis using the Mean Shift algorithm, in combination with some classification abilities based on Learning Vector Quantization (WEKA) algorithm. Far from being commercialized, this thesis suggests a simple way to increase the synergy of three different software packages, in order to set in motion further more solid attempts.

Knowledge from a series of different technologies was brought together to achieve the present result. From Computer Graphic, to Advanced Statistics for Image Analysis and Machine Learning, to GUI development in Java, to Win32 GDI, to JNI and on.

The convergence of the mean shift procedure to the closest mode of the underlying distribution is proven both for the Epanechnikov kernel and the general case of kernels with convex and monotonic decreasing profile. Exploiting parallel mean shift processes a robust clustering method, which was used for the analysis of complex feature spaces derived from RGB satellite data. The cluster centers are obtained by finding the modes of the underlying distribution and their basins of attraction define the cluster boundaries. Examples of image segmentation in a variety of implementations are presented to show the superior performance of the algorithm, as well as the effortless way to scientifically accommodate a wide spectrum of image analysis needs, from a simple user who wants to distinguish the water from the land, or to polygonize all the buildings in his village up to an experienced user who tries to estimate the green spaces in an urban area, or to calculate the burnt area of a forest.

Ευχαριστίες

Αυτό αποδείχθηκε ένα από τα δυσκολότερα κομμάτια της μεταπτυχιακής διατριβής μου. Θα αρκεστώ στο να ονομάσω τους ανθρώπους που συνέβαλαν σε αυτήν την εργασία, λαμβάνοντας υπόψη ότι η αναφορά τους δεν επαρκεί για να περιγράψει την ευγνωμοσύνη μου προς αυτούς.

Η συγγραφή της παρούσας μεταπτυχιακής εργασίας, θα ήταν αδύνατη χωρίς την υπομονή και μεθοδική προσέγγιση του επιβλέποντος καθηγητή και Αργιαλά Δημήτριου και του Εργαστηρίου Ψηφιακής Τηλεπισκόπησης, της σχολής Αγρονόμων & Τοπογράφων Μηχανικών στο Ε.Μ.Π.

Αντίστοιχα θα πρέπει να ευχαριστήσω τον Ian Roessle ο οποίος μου παραχώρησε το λογισμικό EarthWatch για περεταίρω ανάπτυξη, καθώς και τον φίλο και συνεργάτη Αντώνιο Κανέλλη, δίχως την βοήθεια του οποίου, θα ήταν αμφίβολη η τύχη του πειράματος.

Τέλος, δεν θα μπορούσα να μην αναφέρω τους γονείς μου Φανούρη και Ηρώ, καθώς και την αδελφή μου Ελεάννα, που για μια ακόμα φορά στάθηκαν δίπλα μου, από την αρχή ως το τέλος σε αυτή μου την προσπάθεια.

- Ευχαριστώ Πολύ

Τεχνίτης Γεώργιος

1 Εισαγωγή

Σκοπός της παρούσας εργασίας είναι να αξιολογήσει τον αλγόριθμο κατάτμησης μέσης μετατόπισης (mean – shift algorithm) σε εφαρμογές της σύγχρονης τηλεπισκόπησης. Παράλληλα, δημιουργήθηκε ένα μικρό πιλοτικό πρόγραμμα ανοικτού κώδικα, το οποίο επιτρέπει σε ένα μέσο χρήστη να εφαρμόσει αλλά και να παραμετροποιήσει τη συγκεκριμένη μέθοδο, πάνω σε ελεύθερα δορυφορικά δεδομένα του διαδικτύου.

Η δομή της εργασίας έχει ως εξής:

Στο δεύτερο κεφάλαιο, παραθέεται ο κυριότερος όγκος μαθηματικού υποβάθρου πάνω στο οποίο στηρίζεται η διαδικασία Mean Shift. Εκ πρώτης γίνεται ανάλυση των συνθηκών κάτω από τις οποίες λειτουργεί ο αλγόριθμος, περιγράφοντας τη στατιστική διαδικασία ανίχνευσης κορυφών και η ιδιότητα ομαλής τροχιάς. Ακολουθεί η δυναμική ανάλυση των χαρακτηριστικών χώρων με ιδιαίτερη νύξη στο εύρος αυτής, και τέλος υπάρχει μια σειρά πειραμάτων που έχουν γίνει σε κλασικές εικόνες αξιολόγησης αλγορίθμων κατάτμησης.

Το τρίτο και τέταρτο κεφάλαιο περιλαμβάνουν τον κύριο κορμό της εργασίας, θεωρία και εφαρμογή της μεθοδολογίας.

Πιο συγκεκριμένα στο τρίτο κεφάλαιο, περιγράφεται η ελεύθερη κατανομή συσσώρευσης, ο αλγόριθμος λειτουργίας και η εκτίμηση απόδοσης αυτού ενώ στο τέλος το κεφάλαιο ολοκληρώνεται με την περιγραφή κατάτμησης βασισμένη στα κέντρα συσσωρεύσεων.

Το τέταρτο κεφάλαιο αναλύει το περιβάλλον εργασίας του προγράμματος που δημιουργήθηκε. Γίνεται μια μικρή παρουσίαση κάποιων τεχνικών λεπτομερειών του Google Earth, η χρήση των οποίων επέτρεψε τη λειτουργία του προγράμματος κατάτμησης, ενώ δεδομένου ότι είναι βασισμένο σε ένα ήδη υπάρχον πρόγραμμα ανοικτού κώδικα (Earthwatch – University of California) παραθέεται και μια πλήρη περιγραφή αυτού του λογισμικού. Στη συνέχεια ένα διάγραμμα ροής δεδομένων δίνει μια καλύτερη εικόνα για τον τρόπο λειτουργίας του προγράμματος

Το πέμπτο κεφάλαιο συμπεριλαμβάνει εννιά περιπτώσεις μελέτης, κάθε μια από τις οποίες παρατίθεται με τα αποτελέσματα της εφαρμογή τριών ή τεσσάρων διαφορετικών επιπέδων mean shift κατάτμησης, καθώς και τα αντίστοιχα ιστογράμματα αυτών, επιτρέποντας συγκρίσεις και αξιολογήσεις.

Τα συμπεράσματα από την εμπειρία της εργασίας ακολουθούν στο έκτο κεφάλαιο, μαζί με γενικές παραδοχές, παρατηρήσεις αλλά και σχέδια/προτάσεις για μελλοντικές εργασίες επέκτασης του παρόντος εγχειρήματος.

2 Ανασκόπηση βιβλιογραφίας στην κατάτμηση εικόνας βασισμένη σε Mean-Shift-τμηματοποίηση δεδομένων

2.1 Εισαγωγή

Ο σκοπός της κατάτμησης σε μια εικόνα, είναι να ομαδοποιήσει τα εικονοστοιχεία αυτής σε συνιστώσες περιοχές που χρήζουν περισσότερης προσοχής (ανάλυσης), π.χ. περιοχές που αντιστοιχούν σε ξεχωριστές επιφάνειες, οντότητες, , ή διακριτές περιοχές οντοτήτων. Λογικά μία περιοχή που αποτελεί ενδιαφέρον έρευνας, αναμένεται να έχει ομοιογενή χαρακτηριστικά όπως ένταση, υφή κ.α.

Η κατάτμηση θα μπορούσε να χρησιμοποιηθεί για αναγνώριση αντικειμένων, περίφραξη – οριοθέτηση ή υπολογισμό συγκεκριμένων φαινομένων, συμπίεση εικόνας ή και επεξεργασία, βελτιστοποίηση προσπέλασης ενός αλγορίθμου σε μια βάση δεδομένων εικόνων, κτλ.

Οι εργασίες χαμηλού-επιπέδου υπολογιστικής όρασης είναι αρκετά δύσκολες. Πολύ εύκολα μπορούν να αποκομισθούν λανθασμένα αποτελέσματα από τη στιγμή που οι χρησιμοποιούμενες τεχνικές στηρίζονται σε σωστές «προβλέψεις» (guessing) των τιμών των παραμέτρων ρύθμισης από το χρήστη. Για να βελτιωθεί η επίδοση, η εκτέλεση εργασιών χαμηλού-επιπέδου πρέπει να καθορίζεται από το είδος της εργασίας (task-driven), να υποστηρίζεται δηλαδή από ανεξάρτητες πληροφορίες υψηλού επιπέδου. Αυτή η προσέγγιση ωστόσο απαιτεί ότι πρώτον, το στάδιο χαμηλού-επιπέδου παρέχει μια αρκετά αξιόπιστη αναπαράσταση της εισόδου καθώς και ότι η διαδικασία εξαγωγής χαρακτηριστικών ελέγχεται μόνο από πολύ λίγες παραμέτρους ρύθμισης.

Η ανάλυση που βασίζεται στο χώρο χαρακτηριστικών της εικόνας (feature space) είναι ένα παράδειγμα το οποίο μπορεί να επιτύχει τους στόχους που τέθηκαν παραπάνω. Ο χώρος χαρακτηριστικών είναι μια σχεδίαση της εισόδου η οποία λαμβάνεται μέσα από την επεξεργασία των δεδομένων σε μικρά υποσύνολα κάθε φορά. Για κάθε υποσύνολο, μια παραμετρική αναπαράσταση του χαρακτηριστικού που μας ενδιαφέρει λαμβάνεται και το αποτέλεσμα απεικονίζεται σε ένα σημείο στον πολυδιάστατο χώρο της παραμέτρου. Με την ολοκλήρωση της επεξεργασίας ολόκληρης της εισόδου, σημαντικά χαρακτηριστικά αντιστοιχούν στις πυκνότερες περιοχές του χώρου των χαρακτηριστικών, π.χ. σε κλάσεις (clusters), και ο στόχος της ανάλυσης είναι η απεικόνιση αυτών των κλάσεων. Έτσι, θα μπορούσαμε να χρησιμοποιήσουμε τον ορισμό ότι ανάλυση του χώρου των χαρακτηριστικών είναι η διαδικασία ανάκτησης των κέντρων των περιοχών με μεγάλη πυκνότητα, δηλαδή οι αναπαραστάσεις των σημαντικών χαρακτηριστικών της εικόνας.

Το περιεχόμενο ενός συνεχούς χώρου χαρακτηριστικών μπορεί να μοντελοποιηθεί σαν ένα δείγμα από μια πολυδιάστατη, με πολλές κορυφές (multimodal), κατανομή πιθανότητας. Για πραγματικές εικόνες όμως ο αριθμός των κορυφών (modes) μπορεί να είναι πολύ μεγάλος, της τάξεως των δεκάδων ή εκατοντάδων.

Η φύση του χώρου χαρακτηριστικών εξαρτάται από την εφαρμογή. Τα υποσύνολα που χρησιμοποιούνται στη σχεδίαση μπορούν να αποτελούνται ακόμα και από μοναδικά εικονοστοιχεία, όπως συμβαίνει κατά την αναπαράσταση της εικόνας στο χρωματικό χώρο. Τόσο τα πλεονεκτήματα όσο και τα μειονεκτήματα του παραδείγματος του χώρου χαρακτηριστικών προκύπτουν από τη γενική φύση της αναπαράστασης της εισόδου. Από τη μία πλευρά, όλες οι αποδείξεις της παρουσίας ενός σημαντικού χαρακτηριστικού συγκεντρώνονται μαζί, παρέχοντας μια άριστη αντοχή απέναντι στο θόρυβο ο οποίος μπορεί να καταστήσει τις τοπικές αποφάσεις αναξιόπιστες. Από την άλλη πλευρά, χαρακτηριστικά με μικρότερη συμμετοχή στο χώρο των χαρακτηριστικών μπορεί να μην ανιχνευθούν, ακόμα κι αν είναι αξιοπρόσεκτα για την εργασία η οποία εκτελείται. Αυτό το μειονέκτημα ωστόσο, μπορεί να αποφευχθεί σημαντικά είτε αυξάνοντας το χώρο των χαρακτηριστικών με πρόσθετες (χωρικές) παραμέτρους από το χώρο εισόδου, είτε με τη μετέπειτα επεξεργασία του χώρου της εισόδου η οποία να οδηγείται από τα αποτελέσματα της ανάλυσης του χώρου των χαρακτηριστικών (feature space).

Η ανάλυση του χώρου των χαρακτηριστικών εξαρτάται, καθαρά, από την εφαρμογή. Παρόλο που υπάρχει μια πληθώρα δημοσιευμένων τεχνικών ομαδοποίησης (clustering techniques), οι περισσότερες από αυτές δεν επαρκούν για την ανάλυση χώρων χαρακτηριστικών που αντλούνται από πραγματικά δεδομένα. Μέθοδοι οι οποίες στηρίζονται σε μια a priori γνώση του αριθμού των κλάσεων (clusters) που υπάρχουν (συμπεριλαμβανομένων και αυτών που χρησιμοποιούν βελτιστοποίηση ενός γενικού κριτηρίου για να βρουν αυτόν τον αριθμό), καθώς και μέθοδοι οι οποίες αναμφίβολα υποθέτουν το ίδιο σχήμα (συνήθως ελλειπτικό) για όλες τις κλάσεις του χώρου, δεν είναι ικανές να χειριστούν την πολυπλοκότητα ενός πραγματικού χώρου χαρακτηριστικών. Για να μπορέσει κανείς να αποκαλύψει ένα μεγάλο αριθμό σημαντικών χαρακτηριστικών, το πρόβλημα της ανάλυσης του χώρου των χαρακτηριστικών πρέπει να λυθεί σε γενικότερο πλαίσιο. Στις διεργασίες ανάλυσης μιας εικόνας, τα δεδομένα προς ανάλυση προέρχονται από το πεδίο της εικόνας, δηλαδή τα διανύσματα των χαρακτηριστικών ικανοποιούν επιπλέον, χωρικούς περιορισμούς. Αν και αυτοί οι περιορισμοί χρησιμοποιούνται πράγματι στις σύγχρονες τεχνικές, ο ρόλος τους περιορίζεται κυρίως στην εξουδετέρωση των λαθών στην κατανομή των χαρακτηριστικών που συμβαίνουν κατά τη διάρκεια της ανεξάρτητης ανάλυσης του χώρου των χαρακτηριστικών. Για να είναι εύρωστη και δυναμική (robust), η ανάλυση του χώρου των χαρακτηριστικών πρέπει να εκμεταλλεύεται πλήρως την πληροφορία του πεδίου της εικόνας.

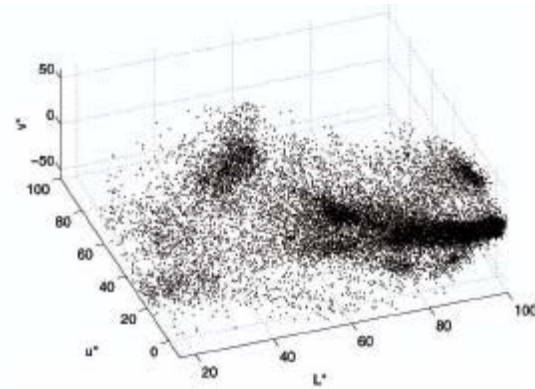
Σαν συνέπεια του αυξημένου ρόλου της πληροφορίας του πεδίου της εικόνας, ο φόρτος στην ανάλυση του χώρου των χαρακτηριστικών μπορεί να μειωθεί. Αρχικά πρέπει να εξαχθούν όλα τα σημαντικά χαρακτηριστικά και μόνο μετά από αυτό προσδιορίζονται οι κλάσεις που περιέχουν τις περιπτώσεις αυτών των χαρακτηριστικών.

Όπως αναφέραμε και παραπάνω, σημαντικά χαρακτηριστικά αντιστοιχούν σε περιοχές μεγάλης πυκνότητας και για να προσδιορίσουμε αυτές τις περιοχές πρέπει να χρησιμοποιηθεί ένα παράθυρο αναζήτησης. Ο αριθμός των παραμέτρων που προσδιορίζουν το σχήμα και το μέγεθος του παραθύρου πρέπει να είναι ελάχιστος κι έτσι, όποτε είναι δυνατόν, ο χώρος χαρακτηριστικών

πρέπει να είναι ισότροπος. Ένας χώρος είναι ισότροπος αν η απόσταση ανάμεσα σε δύο σημεία είναι ανεξάρτητη της θέσης του ζεύγους των σημείων. Στο Σχήμα 2.1 φαίνεται ένα τυπικό παράδειγμα. Η έγχρωμη εικόνα του Σχήματος 2.1a απεικονίζεται στον τρισδιάστατο $L^*u^*v^*$ χρωματικό χώρο.



(a)



(b)

Εικόνα 2.1 - Παράδειγμα του χώρου χαρακτηριστικών. (a) Μια 400 x 276 έγχρωμη εικόνα (b) Ο αντίστοιχος χρωματικός χώρος με 110,400 σημεία πληροφορίας

Αυθαίρετα δομημένοι χώροι χαρακτηριστικών μπορούν να αναλυθούν μόνο με μη-παραμετρικές μεθόδους μια και αυτές δε χρησιμοποιούν υποθέσεις. Αναρίθμητες μη-παραμετρικές μέθοδοι ομαδοποίησης περιγράφονται στη βιβλιογραφία και μπορούν να ταξινομηθούν σε δύο μεγάλες κλάσεις: ιεραρχική ομαδοποίηση και υπολογισμός πυκνότητας. Οι τεχνικές ιεραρχικής ομαδοποίησης είτε συναθροίζουν είτε διαχωρίζουν τα δεδομένα που βασίζονται σε κάποιο μέτρο προσέγγισης. Οι ιεραρχικές μέθοδοι τείνουν να είναι υπολογιστικά ακριβές και ο ορισμός σημαντικού κριτηρίου διακοπής (stopping criterion) για τη συγχώνευση ή διαίρεση των δεδομένων δεν είναι ακριβής.

Η λογική πίσω από τη μη-παραμετρική προσέγγιση ομαδοποίησης που βασίζεται στον υπολογισμό πυκνότητας είναι ότι ο χώρος χαρακτηριστικών μπορεί να θεωρηθεί ως η εμπειρική συνάρτηση πυκνότητας πιθανότητας (p.d.f.) της παραμέτρου που αναπαρίσταται. Έτσι, οι πυκνές περιοχές στο χώρο των χαρακτηριστικών αντιστοιχούν σε τοπικά μέγιστα της p.d.f., δηλαδή στις κορυφές της άγνωστης πυκνότητας. Μόλις καθοριστεί το σημείο της κορυφής, η κλάση που σχετίζεται με αυτό απεικονίζεται βασισμένη στην τοπική δομή του χώρου των χαρακτηριστικών.

Για τον υπολογισμό της πυκνότητας πιθανότητας, υπάρχουν διάφορες μη-παραμετρικές τεχνικές: το πολυδιάστατο ιστόγραμμα, η μέθοδος του «κοντινότερου γείτονα» (nearest neighbor) και ο υπολογισμός με τη χρήση συναρτήσεων πυρήνα. Για χώρους χαρακτηριστικών μεγάλων διαστάσεων, τα πολυδιάστατα ιστογράμματα είναι λιγότερο χρήσιμα λόγω του εκθετικά

αυξανόμενου αριθμού των κουτιών (bins) ανάλογα με τη διάσταση του χώρου. Η μέθοδος του κοντινότερου γείτονα είναι επιρρεπής στον τοπικό θόρυβο (γεγονός που καθιστά την ακριβή ανίχνευση των κορυφών δύσκολη). Για μικρά και μεσαία μεγέθη δεδομένων, ο υπολογισμός με τη χρήση συναρτήσεων πυρήνα είναι μια καλή και πρακτική επιλογή: είναι απλή, και για συναρτήσεις πυρήνα που ικανοποιούν κάποιες συνθήκες, ο υπολογισμός είναι ασυμπτωτικά ακριβής και σταθερός ως προς την έννοια του ελαχίστου τετραγώνου, και ομοιόμορφα κατανεμημένος ως προς την πιθανότητα.

Η προσέγγισή μας για την ανίχνευση των κορυφών και την ομαδοποίηση βασίζεται στη διαδικασία Mean Shift (Μέση Μετατόπιση), η οποία στηρίζεται στην ιδέα της συσχέτισης κάθε σημείου του χώρου των χαρακτηριστικών με όμοιά του, που προτάθηκε το 1975 από τους Fukunaga και Hostetler και «ξεχάστηκε» μέχρι το 1995, όταν ο Cheng με την εργασία του, ανανέωσε το ενδιαφέρον γι' αυτήν. Ο Cheng [CHEN95] γενίκευσε τον αλγόριθμο με τρεις τρόπους και επισήμανε ότι η mean-shift είναι μια διαδικασία εύρεσης κορυφών (mode-seeking) σε μια επιφάνεια που δημιουργείται από τον πυρήνα «σκιά» (shadow kernel). Στο [COME02], ο Comaniciu και ο Meer πρότειναν μια δυναμική προσέγγιση που βασίζεται στο mean-shift, η οποία μπορεί να αναλύσει έναν πολύπλοκο χώρο χαρακτηριστικών με πολλές κορυφές και να απεικονίσει αυστηρά σχηματισμένες κλάσεις. Απέδειξαν τη σύγκλιση μιας αναδρομικής mean-shift διαδικασίας για διακεκριμένα δεδομένα. Παρ' όλες τις άριστες ιδιότητές της, η διαδικασία Mean Shift δεν είναι γνωστή στη στατιστική βιβλιογραφία.

Όπως θα αποδειχθεί στη συνέχεια, μια υπολογιστική μονάδα που βασίζεται στη διαδικασία mean-shift είναι ένα εξαιρετικά ευέλικτο εργαλείο για την ανάλυση του χώρου των χαρακτηριστικών και μπορεί να παρέχει αξιόπιστες λύσεις για πολλές εργασίες της όρασης.

2.2 Η διαδικασία Μέσης Μετατόπισης -Mean-Shift

Μια τεχνική για την αποκάλυψη των σημαντικών χαρακτηριστικών της εικόνας παρουσιάζεται στη συνέχεια. Η τεχνική βασίζεται στον αλγόριθμο mean shift, μια απλή μη-παραμετρική εργασία για τον υπολογισμό των κλίσεων της πυκνότητας. Οι Fukunaga και Hostetler που πρώτοι την εισήγαγαν, στην προσέγγισή τους θεωρούν ότι τα ίδια τα αντικείμενα μπορούν να κινούνται στο χώρο και τελικά να συγκεντρώνονται σε έναν αριθμό κορυφών (tight clumps). Ο αλγόριθμος λειτουργεί με επαναληπτικό τρόπο.

Ο υπολογισμός της πυκνότητας με τη χρήση πυρήνων (kernel density estimation) (γνωστός στη βιβλιογραφία της αναγνώρισης προτύπων και ως τεχνική Parzen παραθύρων) είναι η πιο γνωστή μέθοδος υπολογισμού πυκνότητας. Έχοντας n σημεία δεδομένων $x_i, i=1, \dots, n$ στο d -διάστατο χώρο R^d , ο πολυδιάστατος εκτιμητής της πυκνότητας με τη χρήση πυρήνα $K(x)$ δίνεται από τον τύπο:

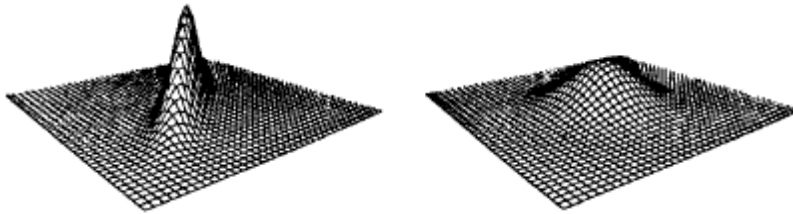
$$1. \hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h^d} K\left(\frac{x-x_i}{h}\right)$$

όπου h είναι η παράμετρος «ομαλοποίησης» ή αλλιώς εύρος (bandwidth). Η συνάρτηση πυρήνα πρέπει να ικανοποιεί τις σχέσεις 2. $K(x) \geq 0$ και 3. $\int K(x)dx = 1$

Μία συνήθως χρησιμοποιούμενη συνάρτηση πυρήνα είναι η πολυδιάστατη κανονική,

$$4. K(x) = (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{1}{2}\|x\|^2\right)$$

η οποία χρησιμοποιήθηκε επίσης στην υλοποίησή μας και στο Σχήμα 2.2 φαίνονται δύο παραδείγματα δυσδιάστατης κανονικής κατανομής με διαφορετική διασπορά.



Εικόνα 2.2. Παραδείγματα δυσδιάστατων κανονικών κατανομών

Ενδιαφερόμαστε μόνο για μια ειδική κατηγορία ακτινικά συμμετρικών πυρήνων που ικανοποιούν τη σχέση

$$5. K(x) = c_{k,d} \cdot k(\|x\|^2)$$

μας και μια συνάρτηση πυρήνα μπορεί να αναπαρασταθεί από τη μορφή του προφίλ της, όπου k είναι η συνάρτηση προφίλ και $c_{k,d}$ μια θετική σταθερά η οποία ικανοποιεί τη σχέση

$$6. c_{k,d} \cdot \int k(\|x\|^2) dx = 1$$

Έτσι, το προφίλ της συνάρτησης πυρήνα Gauss είναι

$$7. k(x) = \exp\left(-\frac{x}{2}\right)$$

και ο αντίστοιχος εκτιμητής της πυκνότητας μπορεί να γραφεί ως εξής:

$$8. \hat{f}_{h,k}(x) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^n k\left(\left\|\frac{x-x_i}{h}\right\|^2\right)$$

Από τη στιγμή που ο d-διάστατος χώρος χαρακτηριστικών μπορεί να θεωρηθεί εμπειρικά, ως η συνάρτηση πυκνότητας πιθανότητας $f(x)$ με τα τοπικά μέγιστα να αντιστοιχούν στις πυκνές περιοχές του χώρου, μπορούμε να βρούμε τα κέντρα κάθε κλάσης ελέγχοντας τα διανύσματα που ικανοποιούν την κλίση $\widehat{\nabla}f(x) = 0$ και κρατώντας μόνο τα τοπικά μέγιστα. Έτσι, το πρώτο βήμα στην ανάλυση του χώρου των χαρακτηριστικών με πυκνότητα $f(x)$ είναι να βρούμε τις κορυφές (modes) αυτής της πυκνότητας και η διαδικασία mean shift αποτελεί έναν «κομψό» τρόπο να προσδιορίσουμε αυτούς τους μηδενισμούς της κλίσης χωρίς να υπολογίσουμε την πυκνότητα.

Αντιστοιχώντας τον υπολογισμό της κλίσης πυκνότητας στην κλίση του υπολογισμού της πυκνότητας, έχουμε:

$$9. \widehat{\nabla}f_{h,K}(x) \equiv \nabla\widehat{f}_{h,K}(x) = \frac{2c_{k,d}}{n \cdot h^{d+2}} \cdot \sum_{i=1}^n \left[(x - x_i) \cdot k' \left(\left\| \frac{x - x_i}{h} \right\|^2 \right) \right]$$

2.2.1 Υπολογισμός της κλίσης της πυκνότητας

Ας ορίσουμε τη συνάρτηση $g(x) = -k'(x)$, υποθέτοντας ότι η παράγωγος του προφίλ της συνάρτησης πυρήνα k υπάρχει για όλα τα $x \in [0, \infty)$. Χρησιμοποιώντας τη $g(x)$ για προφίλ, η συνάρτηση πυρήνα $G(x)$ ορίζεται ως

$$10. G(x) = c_{g,d} g(\|x\|^2)$$

όπου $c_{g,d}$ είναι η αντίστοιχη σταθερά κανονικοποίησης. Η συνάρτηση πυρήνα $K(x)$ καλείται και «σκιά» της $G(x)$. Εδώ πρέπει να τονίσουμε ότι η κανονική συνάρτηση πυρήνα και η «σκιά» της έχουν την ίδια έκφραση. Χρησιμοποιώντας τη $g(x)$, η (3) παίρνει την εξής μορφή:

$$11. \widehat{\nabla}f(x) = \frac{2c_{k,d}}{n \cdot h^{d+2}} \cdot \sum_{i=1}^n g \left(\left\| \frac{x - x_i}{h} \right\|^2 \right) \cdot \left[\frac{\sum_{i=1}^n x_i g \left(\left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^n g \left(\left\| \frac{x - x_i}{h} \right\|^2 \right)} - x \right]$$

Και οι δύο όροι της παραπάνω σχέσης είναι πολύ σημαντικοί. Από τη (8) συμπεραίνουμε ότι ο πρώτος όρος είναι ανάλογος του υπολογισμού της πυκνότητας στο x χρησιμοποιώντας τη συνάρτηση πυρήνα G :

$$12. \widehat{f}_{h,G}(x) = \frac{c_{g,d}}{nh^d} \sum_{i=1}^n g \left(\left\| \frac{x - x_i}{h} \right\|^2 \right)$$

Ο τελευταίος όρος της (11) καλείται μέση μετατόπιση (mean shift), δηλαδή η διαφορά ανάμεσα στο weighted μέσο όρο και το x , το κέντρο της συνάρτησης πυρήνα (παράθυρο), αν θεωρήσουμε τη συνάρτηση g σαν μια συνάρτηση βάρους. Έτσι:

$$13. m_{h,G}(x) = \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x$$

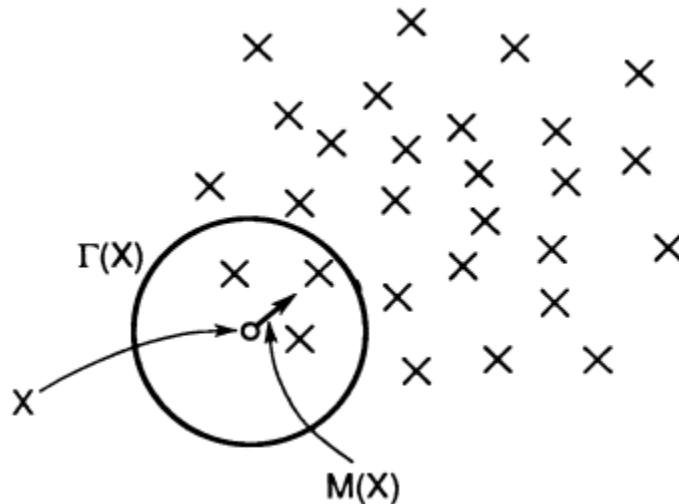
Από τις (12) και (13), η 11 γίνεται:

$$1 \quad \widehat{\nabla} f_{h,K}(x) = \widehat{f}_{h,G}(x) \frac{2c_{k,d}}{h^2 c_{g,d}} m_{h,G}(x)$$

από την οποία προκύπτει η:

$$15. m_{h,G}(x) = \frac{1}{2} h^2 c \frac{\widehat{\nabla} f_{h,K}(x)}{\widehat{f}_{h,G}(x)}$$

Η σχέση (15) δείχνει ότι στη θέση x , το διάνυσμα mean shift υπολογίζεται από τη συνάρτηση πυρήνα G η οποία είναι ανάλογη του υπολογισμού της κανονικοποιημένης κλίσης της συνάρτησης πυκνότητας χρησιμοποιώντας τη συνάρτηση πυρήνα K . Η κανονικοποίηση επιτυγχάνεται με τον υπολογισμό της πυκνότητας στο x χρησιμοποιώντας τη συνάρτηση πυρήνα G . Έτσι, το διάνυσμα mean shift δείχνει πάντα προς την κατεύθυνση της μέγιστης αύξησης στην πυκνότητα. Το Σχήμα 2.3 περιγράφει πώς ο τοπικός μέσος όρος σχετίζεται με την κλίση της συνάρτησης πυκνότητας, όπου $\Gamma(X)$ είναι μια σφαιρική περιοχή, και βλέπουμε ότι ο μέσος όρος των τοπικών δειγμάτων γενικά δείχνει προς την κατεύθυνση της μεγάλης πυκνότητας.

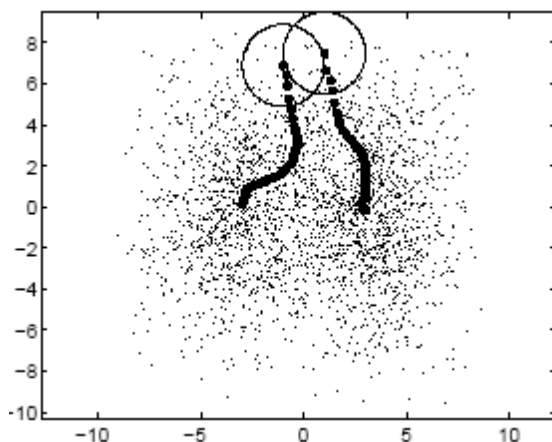


Εικόνα 2.3 Ο τοπικός μέσος όρος ως υπολογισμός της πυκνότητας

Η σχέση (15) είναι «διαισθητική», ο τοπικός μέσος όρος μετατοπίζεται προς την περιοχή στην οποία βρίσκεται η πλειοψηφία των σημείων. Από τη στιγμή που το διάνυσμα mean shift είναι ανάλογο του υπολογισμού της τοπικής κλίσης, μπορεί να ορίσει μια διαδρομή που οδηγεί σε ένα στάσιμο σημείο της υπολογιζόμενης πυκνότητας, όπως φαίνεται στο **Σχήμα 2.4**. Οι κορυφές της πυκνότητας αποτελούν τέτοια στάσιμα σημεία. Η διαδικασία mean shift που επιτυγχάνεται με διαδοχικό

- υπολογισμό του διανύσματος mean shift $m_{h,G}(x)$, και διαδοχική
- «μετάφραση» της συνάρτησης πυρήνα (παράθυρο) $G(x)$ σε $m_{h,G}(x)$,

είναι εγγυημένο ότι θα συγκλίνει σε ένα κοντινό σημείο όπου ο υπολογισμός της σχέσης (8) έχει μηδενική κλίση, όπως θα δειχθεί στην επόμενη ενότητα. Η παρουσία της κανονικοποίησης στον υπολογισμό της πυκνότητας είναι ένα επιθυμητό χαρακτηριστικό. Οι περιοχές με τιμές χαμηλής πυκνότητας δεν προκαλούν κανένα ενδιαφέρον στην ανάλυση του χώρου των χαρακτηριστικών και σε τέτοιες περιοχές τα βήματα του mean shift είναι μεγάλα. Ομοίως, κοντά σε τοπικά μέγιστα τα βήματα είναι μικρά και η ανάλυση πιο ακριβής. Συνεπώς, η διαδικασία mean shift είναι μια προσαρμοστική μέθοδος αναρρίχησης της κλίσης.



Εικόνα 2.4 Διαδοχικοί υπολογισμοί του διανύσματος mean shift προσδιορίζουν μια διαδρομή που οδηγεί στο μέγιστο της τοπικής πυκνότητας

2.2.2 Επαρκής συνθήκη για σύγκλιση

Από τη στιγμή που η g υπολογίζεται αναφερόμενη σε κάθε σημείο δειγματοληψίας x_i , μπορούμε να συμπεράνουμε ότι αναπαριστά τη συνολική δομή όλου του χώρου των χαρακτηριστικών. Τώρα μπορούμε να προσδιορίσουμε τη διαδικασία mean shift ως εξής:

$$16. y_{j+1} = \frac{\sum_{i=1}^n x_i k' \left(\left\| \frac{y_j - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^n k' \left(\left\| \frac{y_j - x_i}{h} \right\|^2 \right)}, j = 1, 2, \dots$$

Ο Comaniciu και ο Meer [5] απέδειξαν ότι αν η συνάρτηση προφίλ k είναι κυρτή και μονοτονικά φθίνουσα, η ακολουθία $\{y_j\}_{j=1,2,\dots}$ θα συγκλίνει σε ένα στάσιμο σημείο. Ως αποτέλεσμα, το διάνυσμα mean shift, το οποίο μπορεί να γραφεί και ως $m(y) = y_{j+1} - y_j$, θα συγκλίνει στο 0, δηλαδή ο υπολογισμός της κλίσης της πυκνότητας θα είναι 0 στο στάσιμο σημείο. Έτσι, μπορούμε εύκολα να βρούμε μια κορυφή του χώρου των χαρακτηριστικών ελέγχοντας αν η αντίστοιχη τιμή πυκνότητας είναι ένα τοπικό μέγιστο. Αυτό αποτελεί μια πιο γενική διατύπωση της ιδιότητας που σημειώθηκε πρώτα από τους Fukunaga και Hostetler [6] και κατόπιν σχολιάστηκε από τον Cheng [7].

Στη σχέση (16), με $\{y_j\}_{j=1,2,\dots}$ σημειώνεται η ακολουθία των διαδοχικών θέσεων του πυρήνα G , όπου, από την (13), y_{j+1} είναι ο weighted μέσος όρος στο y_j που υπολογίζεται με τη χρήση της συνάρτησης πυρήνα G και y_1 είναι το κέντρο της αρχικής θέσης της συνάρτησης πυρήνα. Η αντίστοιχη ακολουθία των υπολογισμών της πυκνότητας με τη χρήση των συναρτήσεων πυρήνα K ,

$$17. \{ \hat{f}_{h,K}(j) \}_{j=1,2,\dots}$$

δίνεται από τη σχέση:

$$18. \hat{f}_{h,K}(j) = \hat{f}_{h,K}(y_j), j = 1, 2, \dots$$

Όπως φανερώνεται από το Θεώρημα 1 που φαίνεται παρακάτω και αποδεικνύεται στο [COME02], μια συνάρτηση πυρήνα K που ικανοποιεί κάποιες συνθήκες αρκεί για τη σύγκλιση των ακολουθιών $\{y_j\}_{j=1,2,\dots}$ και $19. \{y_j\}_{j=1,2,\dots}$ 20. $\{ \hat{f}_{h,K}(j) \}_{j=1,2,\dots}$

Θεώρημα 1. *Αν η συνάρτηση πυρήνα K έχει ένα κυρτό και μονοτονικά φθίνον προφίλ, οι ακολουθίες $\{y_j\}_{j=1,2,\dots}$ και $\{ \hat{f}_{h,K}(j) \}_{j=1,2,\dots}$ συγκλίνουν και η $\{ \hat{f}_{h,K}(j) \}_{j=1,2,\dots}$ είναι μονοτονικά αύξουσα.*

Ωστόσο, όταν κινούμαστε προς την κατεύθυνση της τοπικής κλίσης, η σύγκλιση επιτυγχάνεται μόνο για απειροελάχιστα βήματα. Το μέγεθος του βήματος ενός αλγορίθμου που βασίζεται στην κλίση είναι πολύ σημαντικό για τη συνολική επίδοση. Αν το μέγεθος του βήματος είναι πολύ μεγάλο, ο αλγόριθμος θα αποκλίνει, ενώ αν το μέγεθος του βήματος είναι πολύ μικρό, ο ρυθμός σύγκλισης μπορεί να είναι πολύ αργός. Η εγγυημένη σύγκλιση οφείλεται στην προσαρμοστικότητα του μεγέθους του διανύσματος mean shift, το οποίο επίσης εξαλείφει την ανάγκη για πρόσθετες

διαδικασίες επιλογής των κατάλληλων μεγεθών βημάτων. Αυτό είναι ένα μείζον πλεονέκτημα σε σχέση με τις παραδοσιακές μεθόδους που βασίζονται στις κλίσεις.

Για διακριτά δεδομένα, ο αριθμός των βημάτων για σύγκλιση εξαρτάται από τη χρησιμοποιούμενη συνάρτηση πυρήνα. Όταν G είναι η συνάρτηση πυρήνα ομοιόμορφης κατανομής, η σύγκλιση επιτυγχάνεται σε ένα πεπερασμένο αριθμό βημάτων από τη στιγμή που ο αριθμός των θέσεων που δημιουργούν διακεκριμένες μέσες τιμές είναι πεπερασμένος. Ωστόσο, όταν η συνάρτηση πυρήνα G επιβάλλει μια επιβάρυνση (weighting) στα σημεία δεδομένων μας (ανάλογα με την απόσταση από το κέντρο της), η διαδικασία mean shift συγκλίνει στο άπειρο. Ένας πρακτικός τρόπος για να αποφευχθεί αυτό είναι να θέσουμε ένα κατώτατο όριο για το μέγεθος του mean shift διανύσματος.

2.2.3 Ανίχνευση κορυφών βασισμένη στη διαδικασία Mean Shift

Ας σημειώσουμε με y_c και

$$21. \hat{f}_{h,K}^c = \hat{f}_{h,K}(y_c)$$

τα σημεία σύγκλισης της ακολουθίας

$$22. \{y_j\}_j = 1, 2, \dots$$

και της

$$23. \{\hat{f}_{h,K}(j)\}_{j=1,2,\dots}$$

αντίστοιχα. Οι συνέπειες του Θεωρήματος 1 είναι οι παρακάτω:

- Πρώτον, το διάνυσμα mean shift συγκλίνει στο 0. Πράγματι, από τις (6) και (9), το j -οστό διάνυσμα mean shift είναι:

$$24. m_{h,G}(y_j) = y_{j+1} - y_j$$

και στο όριο,

$$25. m_{h,G}(y_c) = y_c - y_c = 0$$

Με άλλα λόγια, η κλίση του υπολογισμού της πυκνότητας στο σημείο y_c είναι 0

$$26. \nabla \hat{f}_{h,K}(y_c) = 0$$

γεγονός που προκύπτει από την (15). Όμως, το y_c είναι ένα στάσιμο σημείο της,

$$27. \hat{f}_{h,K}$$

- Δεύτερον, από τη στιγμή που η

$$28. \{ \hat{f}_{h,K}(j) \}_{j=1,2,\dots}$$

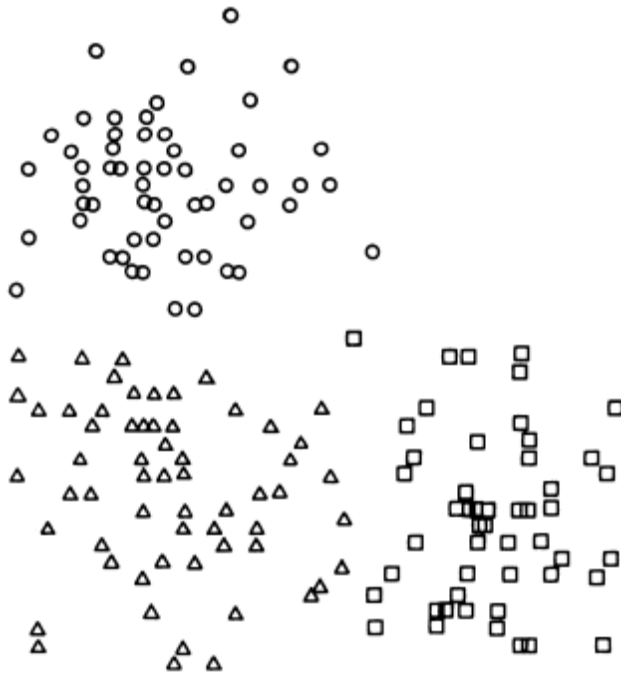
είναι μονοτονικά αύξουσα, οι επαναλήψεις του mean shift ικανοποιούν τις συνθήκες που απαιτούνται από το Θεώρημα «Κατάληψης» (Capture Theorem), το οποίο δηλώνει ότι οι τροχιές τέτοιων μεθόδων κλίσης «έλκονται» από τοπικά μέγιστα αν είναι μοναδικά στάσιμα σημεία (μέσα σε μια μικρή γειτονιά). Μ' άλλα λόγια, καθώς το y_j πλησιάζει αρκετά κοντά σε μια κορυφή της $\hat{f}_{h,K}$, συγκλίνει σε αυτή. Το σύνολο όλων των σημείων που συγκλίνουν στην ίδια κορυφή (mode) ορίζουν μια «λεκάνη έλξης» (basin of attraction) αυτής της κορυφής.

Οι θεωρητικές παρατηρήσεις των παραπάνω συνιστούν έναν πρακτικό αλγόριθμο για ανίχνευση κορυφών:

- Εκτέλεση της διαδικασίας mean shift για να βρεθούν τα στάσιμα σημεία της $\hat{f}_{h,K}$,
- Περιορισμός αυτών των σημείων κρατώντας μόνο τα τοπικά μέγιστα.

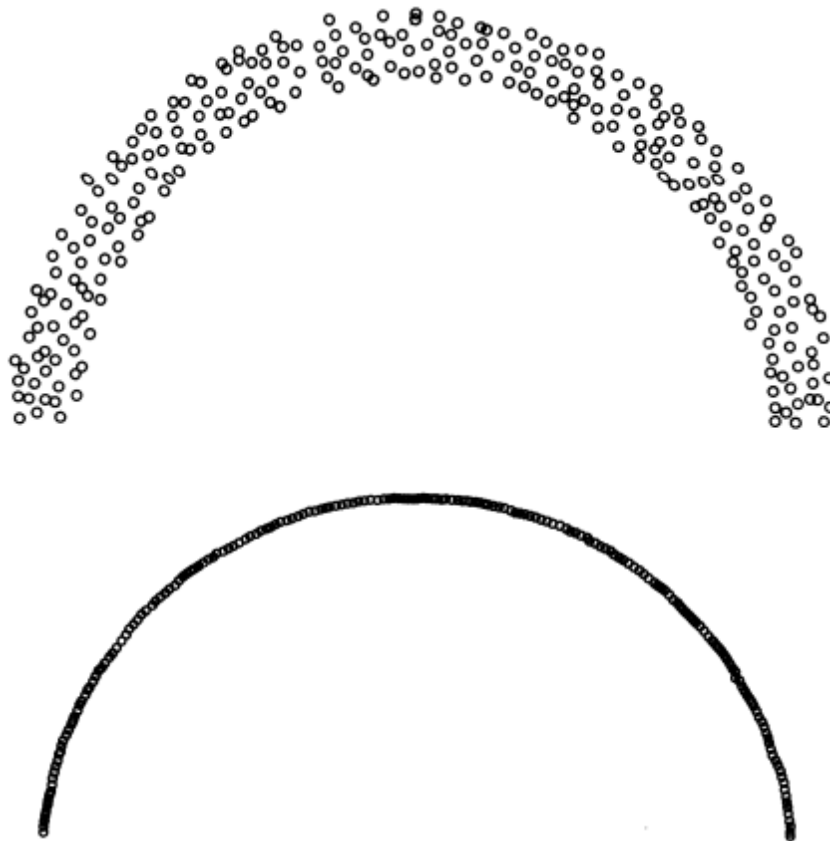
Τα σημεία τοπικών μέγιστων ορίζονται, σύμφωνα με το Θεώρημα «Κατάληψης», ως μοναδικά στάσιμα σημεία μέσα σε μια μικρή ανοιχτή σφαίρα.

Ένα παράδειγμα της εφαρμογής του αλγορίθμου Mean Shift φαίνεται στο Σχήμα 2.5. Τα αρχικά δεδομένα προέκυψαν με τη λήψη 60 παρατηρήσεων από κάθε μια από τις σφαιρικές κανονικές κατανομές με κέντρο τα (0,0), (4,0), και (0,4). Πέντε εφαρμογές της mean shift επανάληψης με $h=1.5$ οδήγησε στη μείωση του συνόλου των δεδομένων σε τρία διακριτά σημεία που βρίσκονται προσεγγιστικά στους μέσους όρους των αρχικών κατανομών. Η αντίστοιχη κατανομή του συνόλου, των αρχικών δεδομένων, φαίνεται στο σχήμα.



Εικόνα 2.5 Ομαδοποίηση τυχαίων δεδομένων με τον αλγόριθμο mean shift

Ο αλγόριθμος μπορεί να χρησιμοποιηθεί επίσης στην αναγνώριση προτύπων για την απομάκρυνση του θορύβου σε εικόνες που αποτελούνται από σημεία. Εδώ το φίλτρο αποτελεί ένα αποτελεσματικό εργαλείο για τον προσδιορισμό της πραγματικής διάστασης των δειγμάτων. Ένα παράδειγμα φαίνεται στο Σχήμα 2.6 το οποίο δείχνει μια διάταξη σημείων πριν και μετά από δύο επαναλήψεις του αλγόριθμου mean shift με την επιλογή κατάλληλων παραμέτρων. Το αποτέλεσμα που προκύπτει ονομάζεται και *υπερεπιφάνεια σκελετού*. Η πρώτη απεικόνιση δείχνει μια κατανομή δειγμάτων τα οποία, αν κριθούν υποκειμενικά, μπορεί να θεωρηθούν είτε μονοδιάστατα είτε δυσδιάστατα. Ας υποθέσουμε λοιπόν ότι η κατανομή είναι μονοδιάστατη και ότι ο ανεπιθύμητος θόρυβος ευθύνεται για τη δυσδιάστατη διασπορά. Για να απομακρύνουμε το θόρυβο, μπορούμε να υπολογίσουμε την κλίση της πυκνότητας σε κάθε σημείο δειγματοληψίας x_i , και το x_i μπορεί να μετακινηθεί προς την κατεύθυνση της κλίσης. Μετά την επανάληψη της διαδικασίας, τα δείγματα εμφανίζονται στην καμπύλη που φαίνεται στο κάτω μέρος του Σχήματος 2.6, η οποία όπως είπαμε ονομάζεται σκελετός της κατανομής. Σύμφωνα με το Fukui[8], η πραγματική διάσταση καθορίζεται από τις τοπικές διαστάσεις, δηλαδή, αν επιλέξουμε μια μικρή περιοχή στη δεύτερη καμπύλη, θα δούμε ότι είναι μονοδιάστατη.



Εικόνα 2.6 Ένα σύνολο δεδομένων πριν και μετά από δύο επαναλήψεις mean shift.

2.2.4 Η ιδιότητα της ομαλής τροχιάς

Η διαδικασία mean shift που χρησιμοποιεί μια κανονική συνάρτηση πυρήνα έχει μια ενδιαφέρουσα ιδιότητα. Η διαδρομή της προς την κορυφή ακολουθεί μια ομαλή τροχιά, με τη γωνία ανάμεσα σε δυο διαδοχικά διανύσματα mean shift να είναι πάντα μικρότερη από 90°.

Χρησιμοποιώντας την κανονική συνάρτηση πυρήνα

$$29. K_N(x) = (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{1}{2}\|x\|^2\right)$$

το j-οστό διάνυσμα mean shift δίνεται από τη σχέση:

$$30. m_{h,N}(y_j) = y_{j+1} - y_j = \frac{\sum_{i=1}^n x_i \exp\left(\left\|\frac{y_j - x_i}{h}\right\|^2\right)}{\sum_{i=1}^n \exp\left(\left\|\frac{y_j - x_i}{h}\right\|^2\right)} - y_j$$

Το ακόλουθο θεώρημα διατηρείται αληθές για όλα τα $j=1,2,\dots$, σύμφωνα με την απόδειξη που δίνεται στο [5].

Θεώρημα 2. Το συνημίτονο της γωνίας που σχηματίζεται ανάμεσα σε δύο διαδοχικά διανύσματα mean shift είναι αυστηρά θετικό όταν χρησιμοποιείται η κανονική συνάρτηση πυρήνα, δηλαδή:

$$31. \frac{m_{h,N}(y_j)^T m_{h,N}(y_{j+1})}{\|m_{h,N}(y_j)\| \|m_{h,N}(y_{j+1})\|} > 0$$

Ως συνέπεια του Θεωρήματος 2, η κανονική συνάρτηση πυρήνα φαίνεται να είναι η ιδανικότερη για τη διαδικασία mean shift. Η ομαλή τροχιά της διαδικασίας mean shift έρχεται σε αντίθεση με τη συνήθη μέθοδο «απότομης ανόδου» (εκτίμηση της τοπικής κλίσης η οποία ακολουθείται από γραμμική μεγιστοποίηση) της οποίας ο ρυθμός σύγκλισης σε επιφάνειες με βαθιές και στενές κοιλάδες είναι μικρός, γεγονός που οφείλεται στην «ζιγκ-ζαγκ» τροχιά της.

Στην πράξη, η σύγκλιση της διαδικασίας mean shift που βασίζεται στην κανονική συνάρτηση πυρήνα απαιτεί μεγάλο αριθμό βημάτων, όπως αναφέρθηκε και στο τέλος της ενότητας 2.2.2. Ωστόσο, η ποιότητα των αποτελεσμάτων είναι πάντα καλύτερη από το αν χρησιμοποιούσαμε για παράδειγμα, την ομοιόμορφη κατανομή ως συνάρτηση πυρήνα, λόγω του ότι η σύγκλιση της είναι πεπερασμένη.

2.2.5 Γενικές παραδοχές

Γενικά οι παραδοχές που διέπουν τον αλγόριθμο έχουν ως εξής,

Έστω R η εικόνα.

Η κατάτμηση χωρίζει την R σε N διακριτές περιοχές R_1, R_2, \dots, R_N με βάση τον κανόνα κατάτμησης $P(R)$

έτσι ώστε να ισχύουν τα παρακάτω:

$$36. R_i \cap R_j = \emptyset \text{ και } 37. \forall i, j, i \neq j$$

a) Οι περιοχές πρέπει να καλύπτουν όλη την εικόνα:

$$38. R = \bigcup_{i=1}^N R_i$$

b) Κάθε κατάτμηση είναι συνδεδεμένη με περιοχή με βάση έναν προκαθορισμένο κανόνα.

c) Οι περιοχές είναι μεταξύ τους ασύνδετες:

d) Όλα τα pixels μίας περιοχής πρέπει να έχουν τις ίδιες ιδιότητες:

$$39. P(R_i) = TRUE \text{ για } 40. i = 1, 2, \dots, N$$

e) Οι περιοχές είναι διακριτές:

$$41. P(R_i \cup R_j) = FALSE$$

2.2.6 Παράμετροι που θα πρέπει να ληφθούν υπόψη κατά την κατάτμηση.

1) Πόσο μεγάλη ακρίβεια περιμένουμε να έχει η κατάτμηση χωρίς να γίνεται η αναγνώριση αντικειμένων μέσα στην εικόνα.

2) Τι προσδιορίζει την κατάτμηση; Πώς θα εκφράσουμε μαθηματικά αυτό το πρόβλημα?

3) Πώς λύνεται το προσδιοριζόμενο πρόβλημα;

4) Πώς μπορούμε να αξιολογήσουμε τα αποτελέσματα;

2.3 Δυναμική ανάλυση των χαρακτηριστικών χώρων

Η ύπαρξη πολλών κορυφών (multimodality) και οι αυστηρά σχηματοποιημένες κλάσεις είναι οι ιδιότητες που ορίζουν έναν πραγματικό χώρο χαρακτηριστικών. Η ιδιότητα της διαδικασίας mean

shift να κινείται προς την κορυφή του λόφου (peak of the hill) στον οποίο ορίστηκε αρχικά, την καθιστά ιδανική υπολογιστική μέθοδο για να αναλύει τέτοιους χώρους.

Πριν να πραγματοποιηθεί η ανάλυση, δύο σημαντικά θέματα πρέπει να συζητηθούν: η μετρική (metric) του χώρου των χαρακτηριστικών και το σχήμα της συνάρτησης πυρήνα. Η αντιστοίχιση του πεδίου εισόδου σε ένα χώρο χαρακτηριστικών συνήθως συσχετίζει μια μη-Ευκλείδεια μετρική με το χώρο. Το πρόβλημα της αναπαράστασης του χρώματος θα συζητηθεί στην ενότητα 2.4, αλλά η χρησιμοποιούμενη παραμετροποίηση πρέπει να ελέγχεται προσεκτικά.

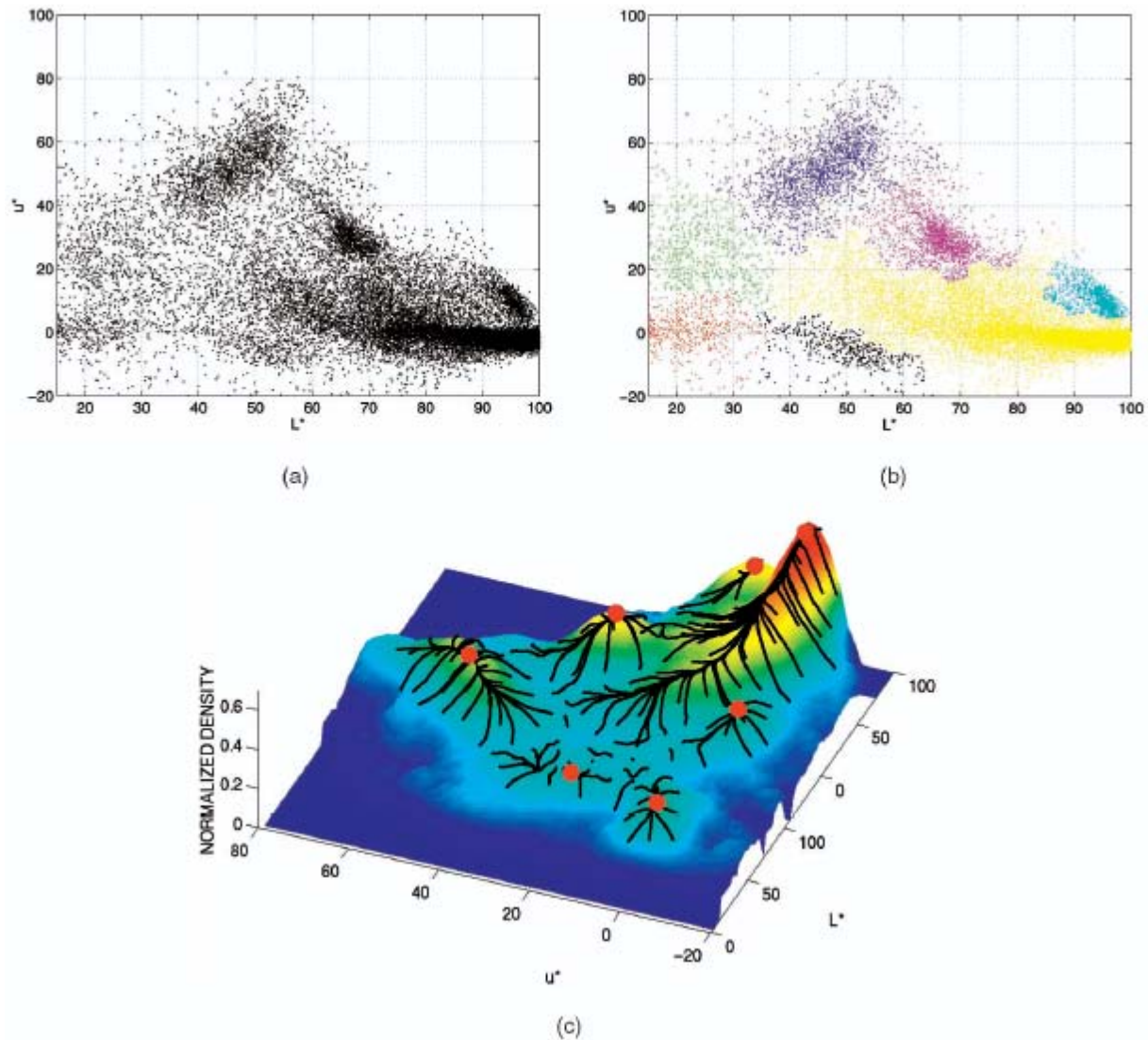
Στην πράξη, είναι προτιμότερο να έχουμε εξασφαλίσει ότι η μετρική του χώρου χαρακτηριστικών είναι Ευκλείδεια έτσι ώστε το εύρος (bandwidth) να ελέγχεται μόνο από μία παράμετρο.

Τα αρχικά σημεία των διαδικασιών mean shift πρέπει να επιλέγονται έτσι ώστε να περικλείουν ολόκληρο το χώρο των χαρακτηριστικών (feature space) (εκτός από τις πολύ αραιές περιοχές). Καθώς τα παράθυρα κινούνται προς τις κορυφές, σχεδόν όλα τα σημεία δεδομένων προσπελούνται και έτσι καθίσταται εκμεταλλεύσιμη όλη η πληροφορία που περιέχεται στο χώρο των χαρακτηριστικών. Πρέπει να σημειώσουμε ότι η σύγκλιση σε μια δεδομένη κορυφή μπορεί να αποκαλύψει ελάχιστα διαφοροποιημένες θέσεις, γεγονός που οφείλεται στο κατώφλι που τερματίζει τις επαναλήψεις. Ομοίως, σε επίπεδες περιοχές κόρου (plateaus), η τιμή της κλίσης είναι σχεδόν μηδενική και η διαδικασία mean shift μπορεί να σταματήσει.

Αυτά τα αποτελέσματα θα μπορούσαν να απαλειφθούν με τη μετέπειτα επεξεργασία (post processing). Τα υποψήφια σημεία σύγκλισης σε απόσταση μικρότερη από το εύρος της συνάρτησης πυρήνα συγχωνεύονται, και αυτό που αντιστοιχεί στη μεγαλύτερη πυκνότητα επιλέγεται.

Η απεικόνιση των κλάσεων είναι ένα φυσικό αποτέλεσμα της διαδικασίας εύρεσης σημείων σύγκλισης. Μετά τη σύγκλιση, η λεκάνη έλξης ενός σημείου σύγκλισης (basin of attraction of a mode), δηλαδή τα σημεία δεδομένων που προσπελάστηκαν από όλες τις διαδικασίες mean shift που συγκλίνουν σε αυτό, αυτόματα απεικονίζουν μια αυστηρά σχηματοποιημένη κλάση.

Η συμπεριφορά αυτής της τεχνικής ανάλυσης του χώρου των χαρακτηριστικών φαίνεται στο Σχήμα 2.7. Ένα δυσδιάστατο σύνολο δεδομένων 110.400 σημείων αναλύεται σε επτά κλάσεις που αναπαρίστανται με διαφορετικά χρώματα (Σχήμα 2.7b). Χρησιμοποιήθηκαν 159 διαδικασίες mean shift με ομοιόμορφη κατανομή. Οι τροχιές τους φαίνονται στο Σχήμα 2.7c, όπου επικαλύπτουν τον υπολογισμό της πυκνότητας (density estimate) που έγινε με τη χρήση Epanechnikov συναρτήσεων πυρήνα. Ο περιορισμός των υποψήφιων σημείων σύγκλισης έδωσε επτά κορυφές. Παρατηρούμε επίσης ότι μερικές από τις τροχιές διακόπτονται πρόωρα από τοπικές περιοχές κόρου (local plateaus).



Εικόνα 2.7 Παράδειγμα μιας δυσδιάστατης ανάλυσης του χώρου των χαρακτηριστικών

2.3.1 Επιλογή του εύρους

Με μία ακριβή προσέγγιση μπορούμε να θεωρήσουμε τέσσερις διαφορετικές τεχνικές για επιλογή εύρους.

- Η πρώτη έχει στατιστικά κίνητρα. Το ιδανικό εύρος που σχετίζεται με τον υπολογισμό της πυκνότητας με τη χρήση των συναρτήσεων πυρήνα (σχέση (1)) ορίζεται ως το εύρος που επιτυγχάνει τον καλύτερο συμβιβασμό ανάμεσα στη ροπή και τη διασπορά του εκτιμητή, για όλα τα $\chi \in \mathbb{R}^d$, δηλαδή ελαχιστοποιεί το AMISE (Asymptotic Mean Integrated Squared Error)

$$32. AMISE(h) = \int_{-\infty}^{\infty} E(\hat{f}_h(y) - f(y))^2 dy \cdot$$

Η δεύτερη τεχνική επιλογής εύρους σχετίζεται με τη σταθερότητα της ανάλυσης. Το εύρος ορίζεται ως το κέντρο της μέγιστης περιοχής λειτουργίας στην οποία ο ίδιος αριθμός κλάσεων προκύπτει για τα δεδομένα μας.

- Για την τρίτη τεχνική, το καλύτερο εύρος (bandwidth) μεγιστοποιεί μια συνάρτηση η οποία εκφράζει την ποιότητα της ανάλυσης (για παράδειγμα τον αριθμό των έγκυρων κλάσεων). Η συνάρτηση αυτή ουσιαστικά εκτιμά το διαχωρισμό και τη συνδετικότητα των κλάσεων.
- Τελικά, εφόσον στις περισσότερες περιπτώσεις η ανάλυση εξαρτάται από τη διεργασία για την οποία προορίζεται (task dependent), πληροφορίες για το εύρος της συνάρτησης πυρήνα μπορούν να αντληθούν από το χρήστη από διεργασία ανώτερου επιπέδου.

2.4 Εφαρμογές

Η τεχνική ανάλυσης του χώρου των χαρακτηριστικών που περιγράφηκε στην προηγούμενη ενότητα εξαρτάται από την εφαρμογή και έτσι μπορεί να χρησιμοποιηθεί για την ανάπτυξη αλγορίθμων όρασης σε μια ευρεία γκάμα διεργασιών. Δύο σχετικές εφαρμογές περιγράφονται στη συνέχεια: η ομαλοποίηση που διατηρεί τις ασυνέχειες (discontinuity preserving smoothing) και η κατάτμηση εικόνας (image segmentation), εφαρμογή που χρησιμοποιήθηκε κατά την υλοποίηση της παρούσας διατριβής. Η ευελιξία της ανάλυσης του χώρου των χαρακτηριστικών επιτρέπει τη σχεδίαση αλγορίθμων κατά τους οποίους ο χρήστης ελέγχει την εκτέλεση μέσα από μία μόνο παράμετρο, την ακρίβεια της ανάλυσης (δηλαδή, το εύρος της συνάρτησης πυρήνα). Από τη στιγμή που η παράμετρος ελέγχου έχει καθαρά φυσική σημασία, οι νέοι αλγόριθμοι μπορούν εύκολα να ενσωματωθούν σε συστήματα που εκτελούν πιο πολύπλοκες εργασίες. Επιπλέον, εξίσου οι grayscale και οι έγχρωμες εικόνες υπόκεινται σε επεξεργασία με τον ίδιο αλγόριθμο.

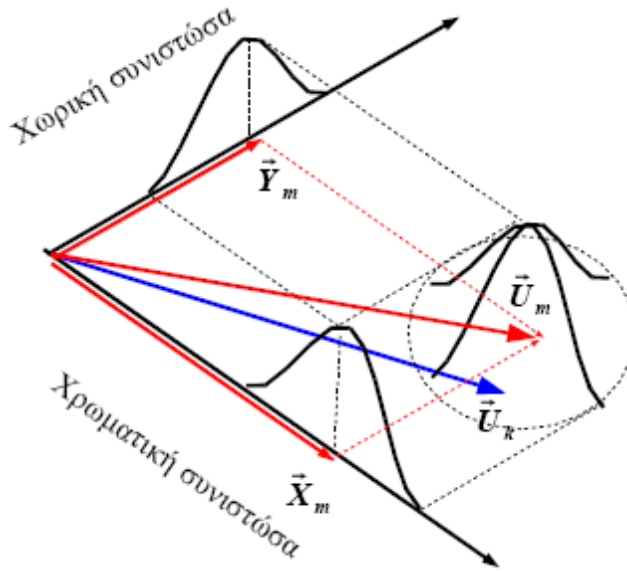
Προτού προχωρήσει κανείς στην ανάπτυξη νέων αλγορίθμων, πρέπει πρώτα να διευθετήσει το θέμα του χρησιμοποιούμενου χρωματικού χώρου. Για να επιτευχθεί μια ουσιώδης κατάτμηση και να εξασφαλισθεί η ισοτροπία του χώρου των χαρακτηριστικών, οι διακρινόμενες διαφορές χρωμάτων πρέπει να αντιστοιχούν σε ευκλείδειες αποστάσεις στο χρωματικό χώρο που επιλέχθηκε να αναπαραστήσει τα χαρακτηριστικά (pixels). Οι χώροι $L^*u^*v^*$ και $L^*a^*b^*$ σχεδιάστηκαν ειδικά για να προσεγγίσουν, όσο το δυνατόν καλύτερα, μοναδικούς, σε επίπεδο αντίληψης, χρωματικούς χώρους. Και στις δύο περιπτώσεις, το L^* , η συνιστώσα της φωτεινότητας (lightness), ορίζεται με τον ίδιο τρόπο, οι δύο χώροι διαφέρουν μόνο στις συνιστώσες χρωματικότητας (chromaticity). Η σχέση και των τριών συνιστωσών με τις «παραδοσιακές» RGB τιμές είναι μη-γραμμική. Στην πράξη, δεν υπάρχει κάποιο ξεκάθαρο πλεονέκτημα του ενός έναντι του άλλου χρωματικού χώρου, από αυτούς που αναφέραμε παραπάνω, στην υλοποίησή μας όμως χρησιμοποιήσαμε τον $L^*u^*v^*$ παρακινούμενοι από την ιδιότητα της γραμμικής απεικόνισης που διαθέτει.

Το παράδειγμα που είδαμε στο Σχήμα 2.7 ήταν μια άμεση εφαρμογή της τεχνικής της ανάλυσης του χώρου των χαρακτηριστικών σε μία $L^*u^*v^*$ αναπαράσταση της έγχρωμης εικόνας. Ωστόσο, από τη στιγμή που η ανάλυση του χώρου των χαρακτηριστικών μπορεί να εφαρμοστεί όπως είναι και σε χώρους περισσότερων διαστάσεων, ενσωματώσαμε επίσης και τις χωρικές συνιστώσες ενός εικονοστοιχείου στην αναπαράστασή του στο χώρο των χαρακτηριστικών. Η αναπαράσταση διασυνδεδεμένου πεδίου (joint domain representation) χρησιμοποιείται στους δύο αλγόριθμους που θα περιγράψουμε παρακάτω και χρησιμοποιήθηκε και στην υλοποίησή μας.

Μια εικόνα ουσιαστικά αναπαρίσταται από ένα δυσδιάστατο πλέγμα p -διάστατων διανυσμάτων (pixels), όπου $p=1$ για την περίπτωση grayscale, $p=3$ για έγχρωμες εικόνες και $p>3$ για την πολυφασματική περίπτωση. Ο χώρος του πλέγματος είναι γνωστός ως χωρικό πεδίο (spatial domain), ενώ η γκρι επιπέδου, έγχρωμη ή φασματική πληροφορία αναπαρίσταται στο πεδίο τιμών (range domain) ή πεδίο χρωμάτων, όπως θα το αποκαλούμε από εδώ και στο εξής για χάριν απλότητας. Και για τα δύο πεδία προϋποθέτουμε την Ευκλείδεια μετρική. Όταν η θέση και τα διανύσματα του πεδίου χρωμάτων (range domain) συνδέονται στο διασυνδεδεμένο χωρικό-χρωματικό πεδίο (joint spatial-range domain) διάστασης $d=p+2$, η διαφορετική φύση τους πρέπει να συνδεθεί με την κατάλληλη κανονικοποίηση. Έτσι, η πολυδιάστατη συνάρτηση πυρήνα ορίζεται ως το γινόμενο δύο ακτινικά συμμετρικών συναρτήσεων πυρήνα και η Ευκλείδεια μετρική επιτρέπει μια μοναδική παράμετρο εύρους για κάθε πεδίο:

$$33. K_{h_s, h_r}(x) = \frac{C}{h_s^2 h_r^p} k\left(\left\|\frac{x^s}{h_s}\right\|^2\right) k\left(\left\|\frac{x^r}{h_r}\right\|^2\right)$$

Όπου x^s είναι το χωρικό μέρος, x^r το χρωματικό μέρος του εικονοστοιχείου, $k(x)$ το κοινό προφίλ που χρησιμοποιείται από τα δύο πεδία, h_s και h_r τα εύρη της συνάρτησης πυρήνα που χρησιμοποιείται, και C η αντίστοιχη σταθερά κανονικοποίησης. Στην πράξη, μια συνάρτηση πυρήνα Epanechnikov ή η κανονική κατανομή πάντα έχουν ως αποτέλεσμα μια ικανοποιητική επίδοση, έτσι ο χρήστης χρειάζεται μόνο να ορίσει την παράμετρο εύρους $h=(h_s, h_r)$, η οποία, ρυθμίζοντας το μέγεθος της συνάρτησης πυρήνα, καθορίζει την ακρίβεια (resolution) της ανίχνευσης κορυφών. Το Σχήμα 2.8 επεξηγεί γραφικά αυτά που περιγράφηκε παραπάνω.



Εικόνα 2.8 Το διασυνδεδεμένο χωρικό-χρωματικό πεδίο και οι συνιστώσες του για το διάνυσμα U

2.4.1 Ομαλοποίηση που διατηρεί την ασυνέχεια (ακμές)

Η ομαλοποίηση με την αντικατάσταση του εικονοστοιχείου στο κέντρο του παραθύρου από το (weighted) μέσο όρο των εικονοστοιχείων στο παράθυρο αδιάκριτα θολώνει την εικόνα, απομακρύνοντας όχι μόνο το θόρυβο αλλά και σημαντική πληροφορία. Οι τεχνικές ομαλοποίησης που διατηρούν την ασυνέχεια, από την άλλη πλευρά, μειώνουν προσαρμοστικά το μέγεθος της ομαλοποίησης κοντά σε απότομες αλλαγές στην τοπική δομή, π.χ. στις ακμές.

Υπάρχει μια μεγάλη ποικιλία προσεγγίσεων για να επιτευχθεί αυτός ο σκοπός, από το προσαρμοστικό φιλτράρισμα του Wiener μέχρι την υλοποίηση ισότροπων και ανισότροπων διαδικασιών τοπικής διάχυσης. Μια μη επαναληπτική τεχνική ομαλοποίησης που διατηρεί τις ασυνέχειες προτάθηκε τελευταία, το αμφίπλευρο φιλτράρισμα. Τα αμφίπλευρα φίλτρα λειτουργούν επίσης στο διασυνδεδεμένο χωρικό-χρωματικό πεδίο. Τα δεδομένα «ζυγίζονται» ανεξάρτητα στα δύο πεδία και το κεντρικό εικονοστοιχείο υπολογίζεται από το (weighted) μέσο όρο του παραθύρου. Η θεμελιώδης διαφορά ανάμεσα στο αμφίπλευρο φιλτράρισμα και τον αλγόριθμο ομαλοποίησης που βασίζεται στο mean shift είναι στη χρήση της τοπικής πληροφορίας.

2.4.1.1 Φιλτράρισμα Mean Shift

Ας υποθέσουμε ότι x_i και z_i , $i=1, \dots, n$ είναι η d -διάστατη είσοδος και τα φιλτραρισμένα εικονοστοιχεία στο διασυνδεδεμένο χωρικό-χρωματικό χώρο αντίστοιχα. Για κάθε εικονοστοιχείο:

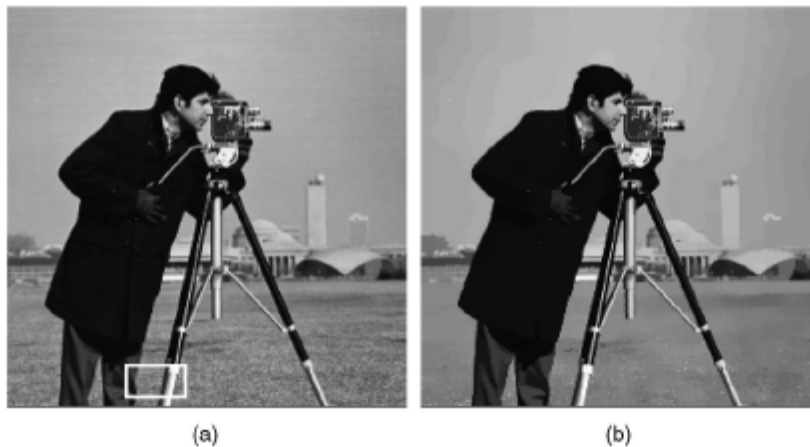
1. Αρχικοποιούμε τα $j = 1$ και $y_{i,1} = x_i$.
2. Υπολογίζουμε το $y_{i,j+1}$ σύμφωνα με την (16) μέχρι να γίνει σύγκλιση, κατά την οποία $y = y_{i,c}$.
3. Θέτουμε (\cdot) .

$$34. Z_i = (x_i^s, y_{i,c}^r)$$

Οι δείκτες s και r δηλώνουν τη χωρική και τη χρωματική συνιστώσα ενός διανύσματος, αντίστοιχα. Η ανάθεση στο βήμα 3 εξασφαλίζει ότι τα φιλτραρισμένα δεδομένα στη χωρική θέση x_i^s θα έχουν τη χρωματική συνιστώσα του σημείου σύγκλισης και μια φυσική ερμηνεία αυτού του γεγονότος δείχνει ότι η έξοδος του mean shift φίλτρου για ένα εικονοστοιχείο ορίζεται η πληροφορία του χρώματος που φέρει το σημείο σύγκλισης. $y_{i,c}^r$

Η συνάρτηση πυρήνα (παράθυρο) στη διαδικασία mean shift κινείται προς την κατεύθυνση της μέγιστης αύξησης στη διασυνδεδεμένη κλίση πυκνότητας (joint density gradient), ενώ το αμφίπλευρο φιλτράρισμα χρησιμοποιεί ένα σταθερό, στατικό παράθυρο.

Στη συχνά χρησιμοποιούμενη grayscale εικόνα 256 x 256 του *κάμεραμαν* (Σχήμα 2.9a) εφαρμόστηκε mean shift φιλτράρισμα με τη χρήση της ομοιόμορφης κατανομής για συνάρτηση πυρήνα, και τα αποτελέσματα φαίνονται στο Σχήμα 2.9b. Οι περιοχές με το γρασίδι έχουν εξομαλυνθεί σχεδόν εξ ολοκλήρου, ενώ λεπτομέρειες όπως το τρίποδο και τα κτίρια στο φόντο έχουν διατηρηθεί. m

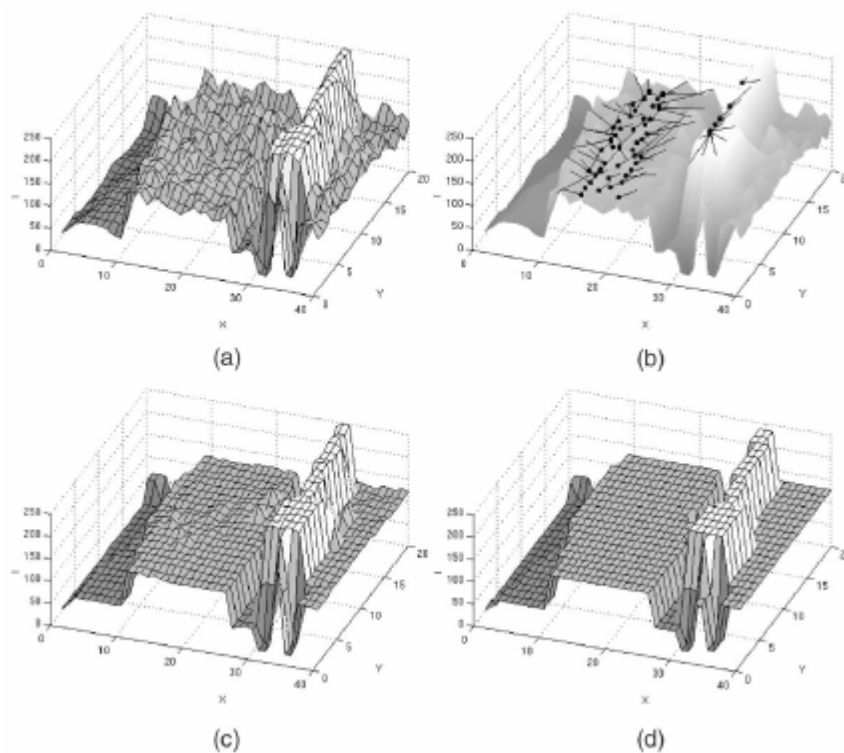


Εικόνα 2.9. Εικόνα του *κάμεραμαν*. (a) Αρχική. (b) Με mean shift φιλτράρισμα $(h_s, h_r) = (8, 4)$.

Για να σχηματίσουμε πιο σαφή εικόνα για τη διαδικασία φιλτραρίσματος, το 40 x 20 παράθυρο που φαίνεται στο Σχήμα 2.9a απεικονίζεται σε τρεις διαστάσεις στο Σχήμα 2.10a. Πρέπει να σημειώσουμε ότι τα δεδομένα απεικονίστηκαν ως προς τον οριζόντιο άξονα για μια πιο κατατοπιστική προβολή. Στο Σχήμα 2.10b φαίνονται τα mean shift μονοπάτια που σχετίζονται με κάθε εικονοστοιχείο (και στις δύο συνιστώσες) της μεγάλης περιοχής (plateau) και της γραμμής. Αξίζει να σημειώσουμε ότι τα σημεία σύγκλισης (μαύρες κουκκίδες) βρίσκονται στο κέντρο της μεγάλης περιοχής (plateau), μακριά από τις ασυνέχειες που τη διακρίνουν. Ομοίως, οι mean shift

τροχιές στη γραμμή παραμένουν πάνω σε αυτή. Σαν αποτέλεσμα, τα φιλτραρισμένα δεδομένα (Σχήμα 2.10c) δείχνουν «καθαρές» ημι-ομογενείς περιοχές.

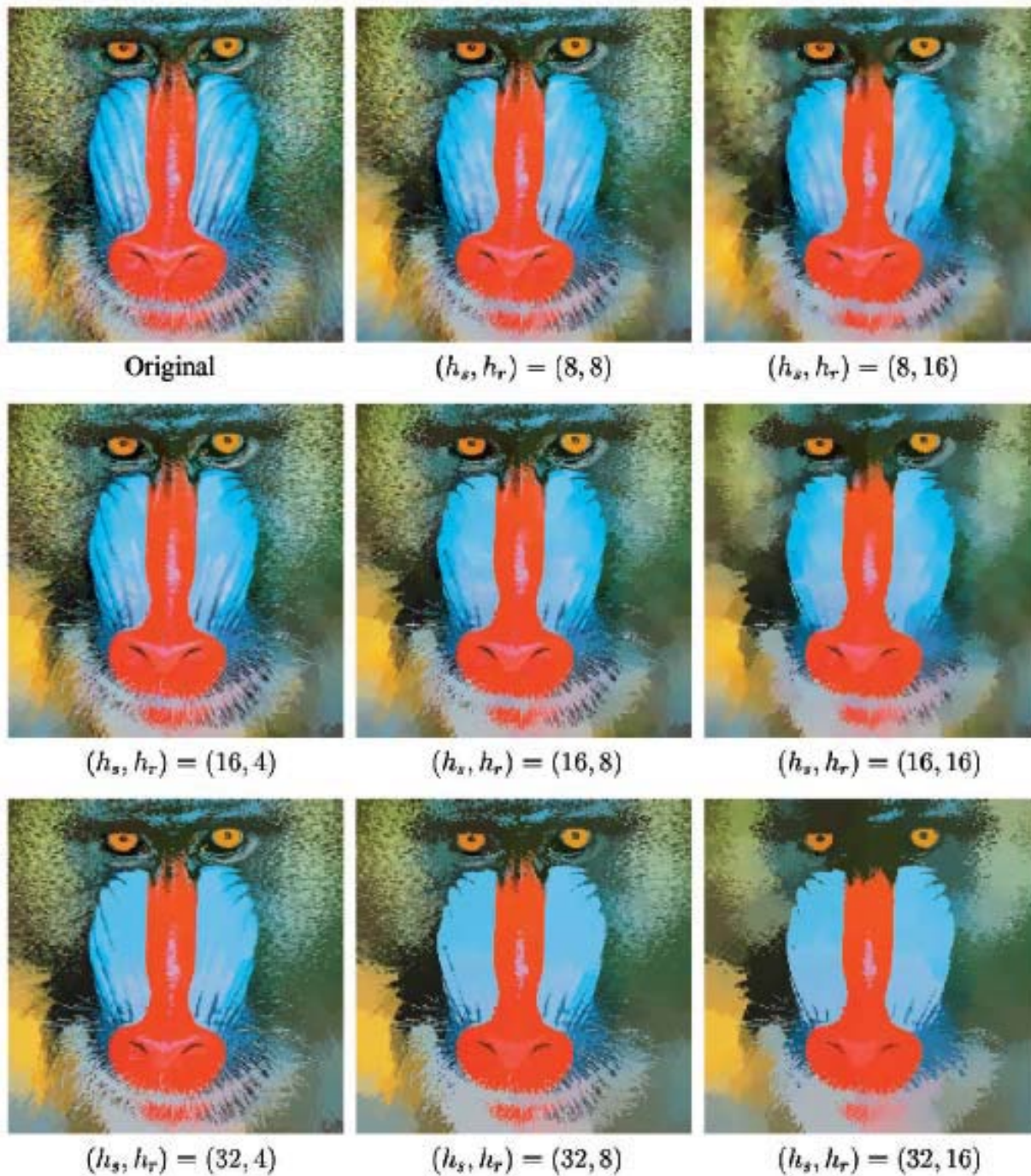
Η φυσική ερμηνεία του βασισμένου σε mean shift φιλτραρίσματος μπορεί εύκολα να αποδοθεί εξετάζοντας το Σχήμα 2.10a, το οποίο, στην πραγματικότητα, δείχνει τις τρεις διαστάσεις του διασυνδεδεμένου πεδίου μιας grayscale εικόνας. Ας πάρουμε ένα εικονοστοιχείο (pixel) στη γραμμή. Η συνάρτηση πυρήνα ομοιόμορφης κατανομής ορίζει ένα παραλληλεπίπεδο με κέντρο αυτό το εικονοστοιχείο και ο υπολογισμός του διανύσματος mean shift παίρνει υπόψη του μόνο αυτά τα εικονοστοιχεία τα οποία έχουν μέσα σε αυτό το παραλληλεπίπεδο και τις χωρικές συνιστώσες τους και τις grayscale τιμές τους. Έτσι, αν το παραλληλεπίπεδο δεν είναι πολύ μεγάλο, υπολογίζεται ο μέσος όρος μόνο εικονοστοιχείων της γραμμής και η νέα θέση του παραθύρου είναι σίγουρο ότι θα παραμείνει πάνω σε αυτή.



Εικόνα 2.10. Mean Shift φιλτράρισμα και κατάτμηση grayscale εικόνας. (a) Είσοδος. (b) Μονοπάτια Mean Shift για τα εικονοστοιχεία της μεγάλης περιοχής και της γραμμής. (c) Αποτέλεσμα φιλτραρίσματος με $(h_s, h_r) = (8, 4)$. (d) Αποτέλεσμα τμηματοποίησης.

Ένα δεύτερο παράδειγμα φιλτραρίσματος φαίνεται στο Σχήμα 2.11. Η 512 x 512 έγχρωμη εικόνα του μπαμπούνου υπέστη επεξεργασία από mean shift φίλτρα τα οποία χρησιμοποίησαν κανονικής κατανομής συναρτήσεις πυρήνα οι οποίες ορίστηκαν από διάφορες χωρικές και χρωματικές αναλύσεις, $(h_s, h_r) = (8 \div 32, 4 \div 16)$ Παρόλο που η υφή του τριχώματος έχει εξαλειφθεί, οι λεπτομέρειες από τα μάτια και τα μουστάκια παραμένουν σαφείς (μέχρι κάποια συγκεκριμένη ανάλυση). Κανείς μπορεί να διαπιστώσει ότι το χωρικό εύρος έχει μια αισθητή επίδραση στην

έξοδο σε σχέση με το χρωματικό εύρος (color bandwidth). Όταν το h_s αυξάνεται τότε μόνο χαρακτηριστικά με μεγάλη χωρική συμμετοχή απεικονίζονται στη φιλτραρισμένη εικόνα. Από την άλλη πλευρά, όσο το h_r μεγαλώνει, μόνο χαρακτηριστικά με υψηλή χρωματική αντίθεση «επιβιώνουν».



Εικόνα 2.11.Εικόνα του μπαμπούνου. Αρχική και φιλτραρισμένη.

2.4.2 Κατάτμηση εικόνας

Η κατάτμηση εικόνας, δηλαδή η ανάλυση μιας grayscale ή μιας έγχρωμης εικόνας σε ομογενείς περιοχές, αποτελεί βάσιμα την πιο σημαντική και ταυτόχρονα ενδιαφέρουσα χαμηλού-επιπέδου διεργασία όρασης. Η ομοιογένεια ορίζεται συνήθως ως ομοιότητα στις τιμές των εικονοστοιχείων.

Η κατάτμηση εικόνας που βασίζεται στη διαδικασία mean shift είναι μια άμεση επέκταση του αλγορίθμου ομαλοποίησης που διατηρεί τις ασυνέχειες. Κάθε εικονοστοιχείο σχετίζεται με ένα σημαντικό σημείο σύγκλισης (significant mode) της πυκνότητας του διασυνδεδεμένου πεδίου που βρίσκεται στη γειτονιά του, αφού τα πολύ κοντινά σημεία σύγκλισης περιοριστούν, όπως συμβάνει και στη γενική τεχνική της ανάλυσης των χαρακτηριστικών χώρου (feature space analysis) (Ενότητα 2.2.3).

2.4.2.1 Mean Shift κατάτμηση / συσσωρευση

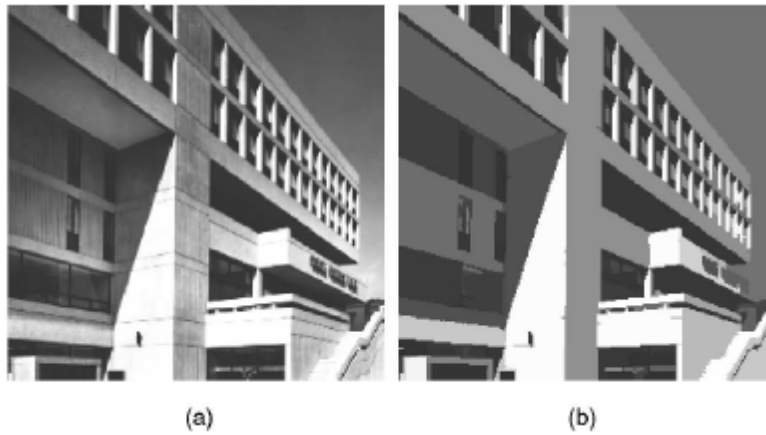
Ας υποθέσουμε ότι x_i και z_i , $i=1, \dots, n$ είναι η d -διάστατη είσοδος και τα φιλτραρισμένα εικονοστοιχεία στο διασυνδεδεμένο χωρικό-χρωματικό χώρο αντίστοιχα και L_i η ετικέτα του i -οστού εικονοστοιχείου της κατατμημένης εικόνας.

1. Εκτελούμε τη mean shift διαδικασία φιλτραρίσματος για την εικόνα και αποθηκεύουμε όλη την πληροφορία για το d -διάστατο σημείο σύγκλισης στο z_i , δηλαδή $z_i = y_{i,c}$.
2. Απεικονίζουμε στο διασυνδεδεμένο πεδίο τις κλάσεις $\{C_p\}_{p=1 \dots m}$ ομαδοποιώντας όλα τα z_i που απέχουν λιγότερο από h_s στο χωρικό πεδίο και h_r στο χρωματικό πεδίο, δηλαδή συνδέουμε τις «λεκάνες έλξης» (basins of attraction) των αντίστοιχων σημείων σύγκλισης.
3. Για κάθε $i=1, \dots, n$, θέτουμε $L_i = \{p \mid z_i \in C_p\}$.
4. Προαιρετικό: Παραλείπουμε τις χωρικές περιοχές οι οποίες περιέχουν λιγότερα από M εικονοστοιχεία.

Το πρώτο βήμα της τμηματοποίησης είναι μια διαδικασία φιλτραρίσματος με τη διαφορά ότι τώρα αποθηκεύεται όλη η πληροφορία για το d -διάστατο σημείο σύγκλισης στο z_i και όχι μόνο το χρωματικό του μέρος.

Το βήμα απεικόνισης των κλάσεων μπορεί να βελτιωθεί σύμφωνα με κάποια a priori πληροφορία και έτσι, κάποιοι άλλοι αλγόριθμοι τμηματοποίησης μπορούν να ενσωματωθούν, όπως για παράδειγμα το νευρωνικό δίκτυο του Kohonen. Το αποτέλεσμα του βήματος της απεικόνισης των κλάσεων φαίνεται στο Σχήμα 2.10d. Ας σημειώσουμε επίσης ότι η συγχώνευση σε μεγαλύτερες περιοχές του αποτελέσματος του φιλτραρίσματος φαίνεται στο Σχήμα 2.10c. Το βήμα της τμηματοποίησης δεν προσθέτει σημαντικό φόρτο στη διαδικασία του φιλτραρίσματος.

Η βελτίωση που οφείλεται στην ανάλυση του διασυνδεδεμένου πεδίου φαίνεται στο Σχήμα 2.12 όπου η 256 x 256 grayscale εικόνα του MIT υπέστη επεξεργασία με $(h_s, h_r, M) = (8, 7, 20)$.



Εικόνα 2.12 Εικόνα του MIT. (a) Αρχική. (b) Κατατετημένη με $(h_s, h_r, M) = (8, 7, 20)$

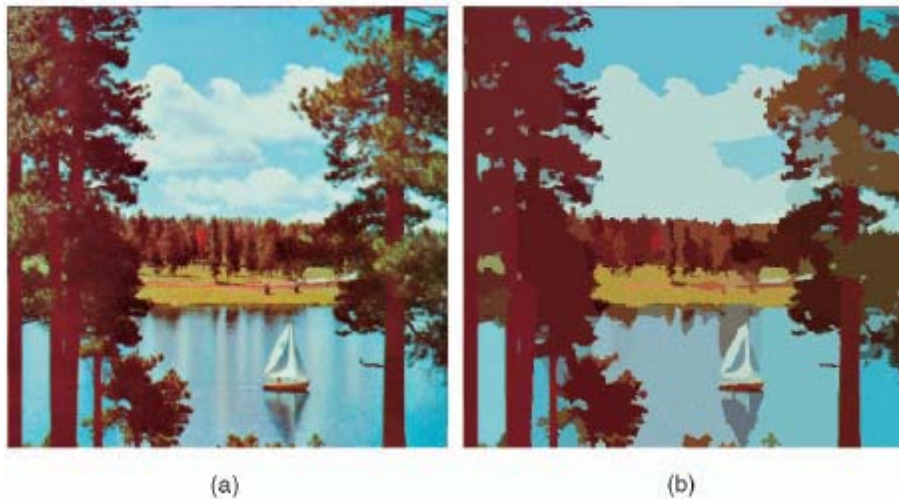
225 ομοιογενείς περιοχές αναγνωρίστηκαν σε κλάσματα του δευτερολέπτου, οι περισσότερες εκ των οποίων απεικονίζουν σημασιολογικά κρίσιμες περιοχές όπως είναι οι τοίχοι, ο ουρανός, τα σκαλοπάτια, η επιγραφή του κτιρίου, κ.ά. Επίσης μπορεί να παρατηρήσει κανείς ότι οι περιοχές

στις οποίες υπάρχει μια μικρή κλίση φωτεινότητας (όπως είναι ο ουρανός) απεικονίστηκαν σαν μια μοναδική περιοχή. Έτσι, η διασυνδεδεμένου πεδίου mean shift κατάτμηση επιτυγχάνει να ξεπερνάει τους εγγενείς περιορισμούς των μεθόδων, βασισμένη μόνο στην grayscale ή έγχρωμη ομαδοποίηση η οποία ουσιαστικά «υπερτμηματοποιεί» περιοχές μικρής κλίσης.

Η κατάτμηση με $(h_s, h_r, M) = (16, 7, 40)$ της 512×512 έγχρωμης εικόνας της λίμνης φαίνεται στο Σχήμα 2.13.

Η παράμετρος του χρώματος h_r και το μικρότερο μέγεθος ενός σημαντικού χαρακτηριστικού M ελέγχουν τον αριθμό των περιοχών στην τμηματοποιημένη εικόνα. Όσο περισσότερο μια εικόνα αποκλίνει από το υποτιθέμενο σταθερό μοντέλο (assumed piecewise constant model), τόσο μεγαλύτερες τιμές πρέπει να χρησιμοποιηθούν για το h_r και το M έτσι ώστε να αποφευχθεί η επίδραση των μικρών τοπικών διαφορών στο χώρο των χαρακτηριστικών. Για παράδειγμα, αν μια εικόνα έχει ένα φόντο με υφή σχηματοποιημένη σε μεγάλο β θμό, τότε πρέπει να χρησιμοποιηθούν μεγάλες τιμές στα h_r και M . Όπως σε κάθε αλγόριθμο όρασης χαμηλού-επιπέδου (low-level vision), η ποιότητα της εξόδου τμηματοποίησης μπορεί να αποτιμηθεί μόνο μέσα στο γενικότερο πλαίσιο όλης της διεργασίας της όρασης και έτσι οι παράμετροι ανάλυσης (resolution parameters) πρέπει να επιλεγθούν σύμφωνα με αυτό το κριτήριο. Ένα σημαντικό πλεονέκτημα της mean shift τμηματοποίησης είναι η συναρμολογησιμότητά της η οποία καθιστά τον έλεγχο της εξόδου τμηματοποίησης πολύ απλό.

Άλλα παραδείγματα τμηματοποίησης στα οποία συγκρίνονται οι αρχικές και οι κατατετμημένες εικόνες φαίνονται στο Σχήμα 2.13.



Εικόνα 2.12 Εικόνα της λίμνης. (a) Αρχική. (b) Τμηματοποιημένη με $(h_s, h_r, M) = (16, 7, 40)$.

3 Ελεύθερη κατανομή συσσώρευσης – θεωρία και εφαρμογή

Το κεφάλαιο αυτό παρουσιάζει έναν πρακτικό αλγόριθμο για την εφαρμογή μη ελεγχόμενης μη-παραμετρικής συσσώρευσης σε μεγάλα σύνολα δεδομένων. Η μέθοδος είναι βασισμένη στην Mean Shift διαδικασία, που είναι απλή, επαρκής, και εύκολη να υλοποιηθεί. Μετά από μια εισαγωγή σε μεθόδους μη-παραμετρικής συσσώρευσης, θα περιγραφεί ο προτεινόμενος αλγόριθμος, ακολουθούμενος από την αξιολόγηση των επιδόσεων του. Τέλος παρουσιάζεται η εφαρμογή του αλγορίθμου στο κομμάτι της χρωματικής (RGB) κατάτμησης.

3.1 Εισαγωγή

Συσχετιζόμενοι με τις δύο εκτιμήσεις (της πυκνότητας και της κλίσης της), υπάρχουν δύο βασικοί αλγόριθμοι της Μη-παραμετρικής συσσώρευσης. Για μια συγκεκριμένη ακτίνα παραθύρου h , και οι δύο αλγόριθμοι αυτόματα εντοπίζουν τον αριθμό των υφιστάμενων συσσωρεύσεων καθώς και των αντίστοιχων ορίων τους.

Χρησιμοποιώντας την εκτίμηση πυκνότητας (2.2), μία ιεραρχική δομή των δεδομένων μπορεί να επιτευχθεί ως ακολούθως.: Για κάθε σημείο x_i , αναζητείται ένα γειτονικό, προοριζόμενο για γονικό στοιχείο x_j , του οποίου η ποσότητα

$$(35) \left[\hat{f}(x_j) - \hat{f}(x_i) \right] \cdot \|x_j - x_i\|^{-1}$$

είναι θετική και μέγιστη, δηλαδή x_j είναι η πιο απότομη ανοδική τιμή σε σχέση με το x_i . Εάν η παραπάνω ποσότητα είναι αρνητική για όλες τις γειτονικές x_j , τότε η x_i χαρακτηρίζεται ως η κομβική ρίζα της δομής του δέντρου. Οι κομβικές ρίζες υποτίθεται ότι είναι κοντά στη λειτουργία της υποκείμενης κατανομής. Η συσσώρευση γίνεται με φυσικό τρόπο ακολουθώντας τις διακλαδώσεις της δομής. Ο αλγόριθμος, που ονομάζεται διάγραμμα θεωρητικής συσσώρευσης, περιγράφεται λεπτομερώς στο [9]. Η ιεραρχική δομή μπορεί επίσης να προέρχεται από επαναληπτικά όρια [10] ή μέσω διαμοιρασμού [11] της εκτίμησης πυκνότητας.

Ο δεύτερος αλγόριθμος χρησιμοποιεί την εκτίμηση της βαθμωτής μεταβολής της πυκνότητας για την ορίσει μια επαναληπτική, «αναρριχητική» τεχνική η οποία ανιχνεύει τις μεθόδους και τις «κοιλιάδες» της υποκείμενης διανομής. Η διαδικασία Mean Shift είναι μια προσαρμοστική τεχνική απότομης ανηφόρας που υπολογίζει το διάλυσμα μέσης μετατόπισης για κάθε σημείο δεδομένων, μετασχηματίζει τον πυρήνα με το βάση τον υπολογισμό αυτό, και επαναλαμβάνει τους υπολογισμούς μέχρι να επιτευχθεί ένα συγκεκριμένο πρότυπο. Γενικεύσεις της Mean Shift

διαδικασίας και εφαρμογές για την συσσώρευση αναφέρονται στο [12]. Μια παραλλαγή της Mean Shift, που ονομάζεται συσσώρευση μέγιστης εντροπίας παρουσιάζεται στο [13]

Ωστόσο, η συσσώρευση μέσω της εφαρμογής της τεχνικής Mean Shift προς κάθε σημείο των δεδομένων δεν μπορεί να είναι ικανοποιητική σε πρακτικές εφαρμογές εφόσον η σύγκλιση σε περιφέρειες με χαμηλή πυκνότητα είναι κακή, ενώ περιοχές με υψηλή πυκνότητα μπορούν να παρουσιάσουν τα λεγόμενα «οροπέδια» πληροφορίας, χωρίς κάποιο σαφή τοπικό μέγιστο (peak)

Όταν το σύνολο των δεδομένων είναι μεγάλο (π.χ. πάνω από 10000 σημεία) το πιο σημαντικό μειονέκτημα των δύο προαναφερόμενων αλγορίθμων είναι η πολυπλοκότητα στους υπολογισμούς τους. Χρειάζονται την πυκνότητα ή την εκτίμηση βαθμωτής μεταβολής της πυκνότητας σε κάθε σημείο δεδομένων που έχει πολυπλοκότητα O για ένα σύνολο n δεδομένων. Το πρόβλημα πολυπλοκότητας επιφέρεται από την έλλειψη της αποδοτικότητας σε αναζήτηση πολυδιάστατου φάσματος (η αναζήτηση για τα σημεία δεδομένων που υπάγονται σε μια συγκεκριμένη γειτονία). Σε αντίθεση με την αναζήτηση της κοντινότερης γειτονίας (nearest neighborhood search) γειτονική που μπορεί να πραγματοποιηθεί σε λογαριθμικό χρόνο [14,15], οι επιδόσεις των d -διαστατών δένδρων που χρησιμοποιούνται στην αναζήτηση ως προς εύρος είναι κάπως δύσκολες για την πρόβλεψη τυχαίων δεδομένων

Στη συνέχεια παρουσιάζουμε έναν πιθανολογικού τύπου Mean Shift αλγόριθμο που θα λαμβάνει υπόψη του τις δυσκολίες που αναφέρθηκαν μέχρι τώρα, του οποίου η πολυπλοκότητα είναι $O(mn)$, με $m \ll n$.

3.2 Ανάλυση Συσσώρευσης

Τα βήματα του αλγορίθμου διατυπώνονται παρακάτω.

1. Καθορισμός μιας τυχαίας ψηφίδωσης (tessellation) του χώρου με $m \ll n$ σφαίρες $S_h(x)$. Για να μειωθεί το υπολογιστικό φορτίο, ένα σύνολο από σημεία m για $x_1 \dots x_m$ που ονομάζεται δείγμα συνόλων επιλέγεται τυχαία από τα δεδομένα. Δύο περιορισμοί επιβάλλονται στα σημεία του δείγματος:
 - η απόσταση μεταξύ δύο οποιωνδήποτε γειτόνων δεν πρέπει να είναι μικρότερη του h , η ακτίνα της σφαίρας $S_h(x)$, και
 - τα δειγματικά σημεία δεν πρέπει να βρίσκονται σε αραιοκατοικημένες περιοχές. Η τελευταία αυτή προϋπόθεση είναι απαραίτητη για να αποφευχθούν ομάδες χαμηλής πυκνότητας. Η περιοχή είναι αραιοκατοικημένη, όταν ο αριθμός των

σημείων στο εσωτερικό της σφαίρας είναι μικρότερος από το κατώφλι T . Οι περιορισμοί απόστασης και πυκνότητας καθορίζουν αυτόματα το μέγεθος m του δειγματικού συνόλου. Οι σφαίρες στο κέντρο του δειγματικού συνόλου καλύπτουν τα περισσότερα από τα σημεία δεδομένων. Όταν ο χρόνος επεξεργασίας δεν είναι τόσο κρίσιμος, ο περιορισμός απόστασης μπορεί να είναι πιο χαλαρός, με αποτέλεσμα να αυξάνεται η ανάλυση της ψηφίδωσης.

2. Εφαρμογή της Mean Shift ανάλυσης στο δειγματικό σύνολο. Ένα σύνολο που περιέχει m υποψήφιους κέντρου συσσώρευσης καθορίζεται από τα σημεία σύγκλισης των m Mean Shift διαδικασιών. Αξιοσημείωτη η μείωση της υπολογιστικής πολυπλοκότητας που τώρα είναι $O(mn)$, με $m \ll n$, και ότι οι υπολογισμοί του Mean Shift διανύσματος βασίζονται σχεδόν σε ολόκληρο το σύνολο των δεδομένων. Έτσι, η ποιότητα της εκτίμησης της βαθμωτής μεταβολής της πυκνότητας δεν μειώνεται από τη χρήση της δειγματοληψίας.
3. Αλλαγή των υποψήφιων κέντρου συσσώρευσης και εφαρμογή ξανά της διαδικασίας Mean Shift. Δεδομένου ότι ένα τοπικό οροπέδιο μπορεί να διακόψει πρόωρα τις επαναλήψεις, κάθε υποψήφιο κέντρο συσσώρευσης δύναται να αλλάξει από ένα τυχαίο διάνυσμα ενός μικρού ελεγκτικού κανόνα και να επιτρέψει στη Mean Shift διαδικασία να συγκλίνει και πάλι.
4. Άντληση των κέντρων συσσώρευσης $\gamma_1 \dots \gamma_p$ από τους υποψήφιους κέντρου συσσώρευσης. Κάθε υποσύνολο των υποψηφίων κέντρου συσσώρευσης που βρίσκονται αρκετά κοντά το ένα στο άλλο (για κάθε δεδομένο σημείο στο υποσύνολο υπάρχει τουλάχιστον ένα άλλο σημείο στο υποσύνολο όπου η απόστασή τους είναι μικρότερη από h), καθορίζει ένα κέντρο συσσώρευσης. Το κέντρο συσσώρευσης είναι ο μέσος όρος των υποψηφίων κέντρου συσσώρευσης στο υποσύνολο. Θα πρέπει να σημειωθεί ότι $p \leq m$.
5. Επικύρωση / Αξιολόγηση των κέντρων συσσώρευσης. Μεταξύ δύο κέντρων συσσώρευσης γ_i, γ_j μία σημαντική στατιστική “κοιλιάδα” θα πρέπει να εμφανιστεί στην υποκείμενη πυκνότητα. Η ύπαρξη της κοιλιάδας ελέγχεται για κάθε ζεύγος (γ_i, γ_j) . Η σφαίρα $S_h(x)$ κινείται με βηματισμό h κατά μήκος της γραμμής που καθορίζεται από την (γ_i, γ_j) και ο σταθμισμένος αριθμός των σημείων δεδομένων που βρίσκονται στη σφαίρα υπολογίζεται σε κάθε θέση, δηλαδή η πυκνότητα υπολογίζεται με χρήση του πυρήνα Epanechnikov (K_E) κατά μήκος της γραμμής. Όταν η αναλογία μεταξύ

$$(36) \min \left[\hat{f}(x_j) \cdot \hat{f}(x_i) \right]$$

και η ελάχιστη πυκνότητα κατά μήκος της γραμμής είναι μεγαλύτερη από ένα κατώφλι T_2 , μία κοιλάδα θεωρείται ότι υπάρχει μεταξύ του y_i και y_j . Αν δεν διαπιστωθεί κάποια κοιλάδα μεταξύ y_i και y_j , το κέντρο διασποράς με τη μικρότερη πυκνότητα (y_i ή y_j) αφαιρείται από το σύνολο των κέντρων διασποράς.

6. Απεικόνιση των συσσωρεύσεων. Στο στάδιο αυτό κάθε δειγματικό σημείο συνδέεται με ένα κέντρο διασποράς. Για την κατανομή των σημειακών δεδομένων χρησιμοποιείται η τεχνική του πλησιέστερου γείτονα (k – nearest neighbor) , δηλαδή κάθε σημείο δεδομένων ανήκει στη συσσώρευση η οποία καθορίζεται από την πλειοψηφία των k -πλησιέστερων σημείων του δείγματος .

3.3 Εκτίμηση απόδοσης

Ο αλγόριθμος συσσώρευσης δεδομένων κάνει χρήση τριών παραμέτρων:

- της αναζήτησης ακτίνας σφαίρας h η οποία ελέγχει την ευαισθησία της ανάλυσης,
- του ορίου T_1 που επιβάλλει τον περιορισμό της πυκνότητας,
- του ορίου T_2 που αντιστοιχεί στο ελάχιστο αποδεκτό ποσοστό αναλογίας στη σχέση κορυφής – κοιλάδας.

Οι παράμετροι T_1 και T_2 έχουν γενικά χαμηλή επιρροή στο αποτέλεσμα. Όλα τα αποτελέσματα πειραμάτων που περιγράφονται εδώ ελήφθησαν με $T_1 = 50$ και $T_2=1.2$. Πέραν των ειδικών περιπτώσεων που υπάρχει σημείωση για διαφορετικές τιμές, , χρησιμοποιήθηκε παντού $k = 1$, δηλαδή οι συσσωρεύσεις οριοθετήθηκαν χρησιμοποιώντας τον κοντινότερο γείτονα στο τελευταίο βήμα του αλγορίθμου. Δεδομένου ότι τα πειραματικά σύνολα δεδομένων είχαν διαφορετικές κλίμακες, η ακτίνα σφαίρας h άλλαξε αντίστοιχα. Σημαντική σημείωση επίσης, ότι η βελτιωμένη Ανισότητα Απόλυτου Σφάλματος [16] χρησιμοποιήθηκε για να υπολογιστούν επαρκώς οι Ευκλείδειες αποστάσεις.

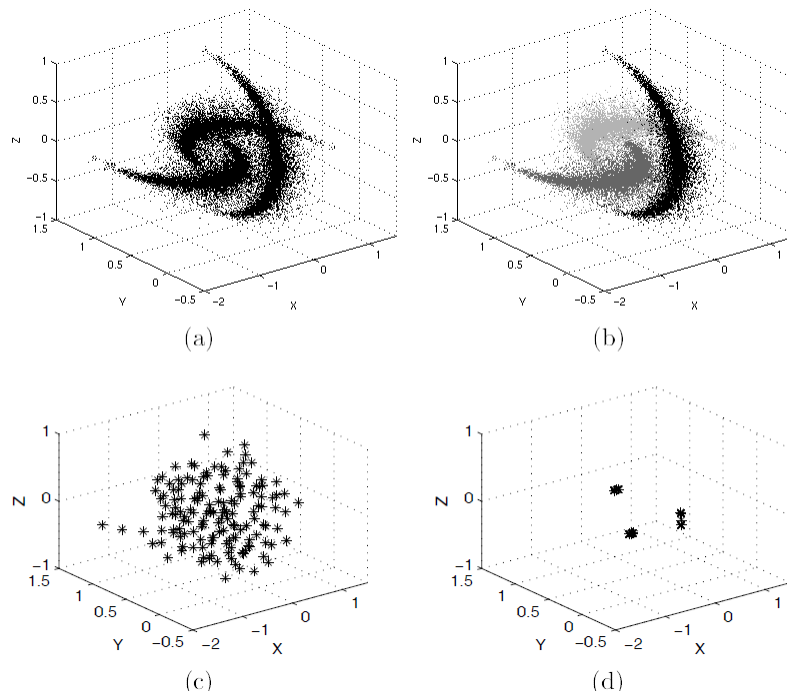
Στη συνέχεια, παρατίθενται 5 πειράματα που επαληθεύουν τις δυνατότητες του αλγορίθμου συσσώρευσης. Θα πρέπει να σημειωθεί ότι δεδομένης της έλλειψης δυνατότητας για πειράματα σε τρισδιάστατα δεδομένα, τα πειράματα 1, 2 και 5 έχουν βασιστεί στα αποτελέσματα του διδακτορικού του Dorin I. Comaniciu, με τίτλο Nonparametric robust methods for computer vision.

Πείραμα 1: Το πρώτο παράδειγμα παρουσιάζεται στην Εικόνα 3.1. Το σύνολο των δεδομένων περιέχει 32460 σημεία με διάσταση $d=3$, ταξινομημένα σε 3 μη γραμμικές διαχωρισμένες συσσωρεύσεις (Εικόνα 3.1α). Μία τυποποιημένη μη-επιβλεπόμενη διαδικασία όπως η ISODATA [17] θα αποτύγχανε σε αυτά τα δεδομένα. Χρησιμοποιώντας μια ακτίνα $h = 0.2$, το ληφθέν σύνολο δειγμάτων είχε 167 σημεία (Εικόνα 3.1c) και συγκλείνανε σε 3 κέντρα διασποράς. Στην εικόνα 3.1b

παρουσιάζονται οι 3 συσσωρεύσεις που προέκυψαν , έχοντας 11050, 10874 και 10716 σημεία, αντίστοιχα.

Πείραμα 2: Ένα απλό παράδειγμα συσώρευσης φαίνεται στο Σχήμα 3.2. Ο σκοπός αυτού του πειράματος ήταν να συγκρίνει τις επιδόσεις του Μη-παραμετρικού αλγορίθμου με την επίδοση του κλασικού Bayes ταξινομητή. Τα σύνολα δεδομένων που περιέχονται σε 10000 σημεία με διαστάσεις $d=3$ προέρχονται από δύο κανονικές κατανομές με συνδιακυμάνσεις 102^{-1} και μέσο διάνυσμα $(0,0,0)T$ και $(40,0,0)T$ αντίστοιχα.

Η εικόνα 3.2b δείχνει τις οριοθετημένες ομάδες που αντιστοιχούν σε μία ακτίνα $h = 10$. Χρησιμοποιώντας τον ταξινομητή Bayes το ποσοστό σφάλματος είναι 2.34%, που οφείλεται στα 234 επικαλυπτόμενα σημεία . Η εικόνα 3.3 δείχνει το ποσοστό σφάλματος που προέκυψε από τον αλγόριθμο για σφαιρικές ακτίνες μεταξύ 8 και 20.

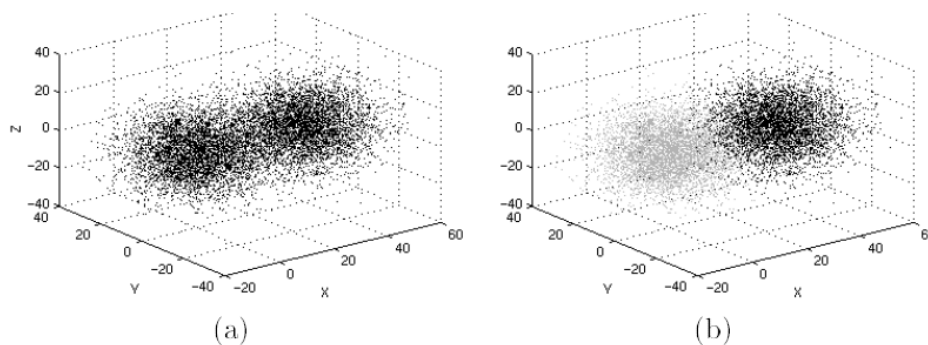


Εικόνα 3.1: Πρώτο πείραμα. (α) Αρχικό σύνολο δεδομένων (32460 σημεία). (β) Απεικόνιση συσσωρεύσεων (κάθε ομάδα εκπροσωπείται με διαφορετική κλίμακα του γκρι) (γ) Σύνολο δειγμάτων (δ) Υποψήφια κέντρα συσώρευσης. Τα σημεία δεδομένων προέρχονται από τυχαία δειγματοληψία και έδωσαν τα σύνολα δειγμάτων που αποτελούν υποθετικά κέντρα συσώρευσης, σημεία αναφοράς για την λειτουργία σύγκλισης της συσώρευσης, από την οποία προκύπτουν και τα τελικά κέντρα συσώρευσης.

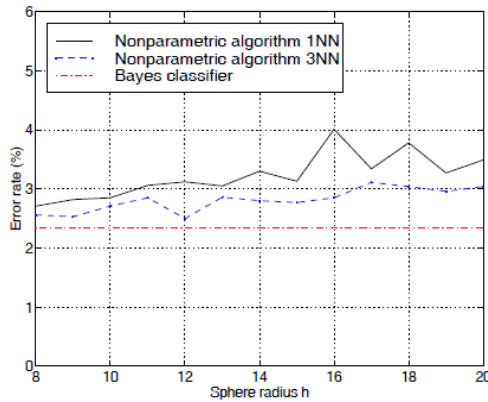
Η κατανομή των σημειακών δεδομένων για τις μεθόδους που χρησιμοποίησαν κοντινότερους γείτονες k , όπου το k ελήφθη ως 1 και 3, αντίστοιχα. Τα ποσοστά σφάλματος με ακτίνα h αυξάνονται λόγω της αύξησης της οριοθέτησης σφάλματος. Την ίδια στιγμή, το ποσοστό σφάλματος είναι χαμηλότερο όταν οι 3 κοντινότεροι γείτονες χρησιμοποιούνται για την κατανομή των σημείων δεδομένων. Η ευθεία γραμμή που εμφανίζεται στο διάγραμμα αντιπροσωπεύει το ποσοστό σφάλματος Bayes. Οι επιδόσεις του Μη-παραμετρικού αλγόριθμου είναι πολύ κοντά με εκείνες του ταξινομητή Bayes, παρά την έλλειψη της αρχικής γνώσης που χρησιμοποιείται στην περίπτωση της πρώτης.

3.4 Κατάτμηση βασισμένη στα κέντρα συσσωρεύσεων

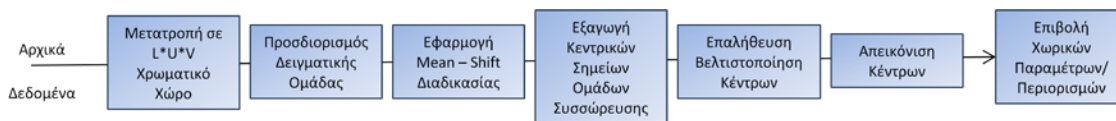
Αυτό το παράρτημα δείχνει την εφαρμογή του νέου αλγόριθμου στην κατάτμηση των έγχρωμων εικόνων. Σύμφωνα με τη συνοπτική μέθοδο που παρουσιάζεται στην εικόνα 3.4, η ομαδοποίηση εκτελείται στον $L*U*V^*$ χρωματικό χώρο όπου κάθε οριοθετημένη συσσώρευση αντιστοιχεί σε ομοιογενείς περιφέρειες της εισαγόμενης εικόνας. Η κατατετμημένη εικόνα προέρχεται από τη χαρτογράφηση του χρωματικού διανύσματος στον χώρο της εικόνας και την επιβολή χωρικών περιορισμών.



Εικόνα 3.2: Δεύτερο πείραμα (α) Αρχικό σύνολο δεδομένων (10000 σημεία). (b) Οριοθέτηση συσσώρευσης (2 ομάδες).



Εικόνα 3.3: Το ποσοστό σφάλματος του προτεινόμενου αλγόριθμου για διαφορετικές τιμές της σφαιρικής ακτίνας και διαφορετικών αριθμών των κοντινότερων γειτονικών στοιχείων. Η διακεκομμένη γραμμή αντιπροσωπεύει το ποσοστό σφάλματος του ταξινομητή Bayes.

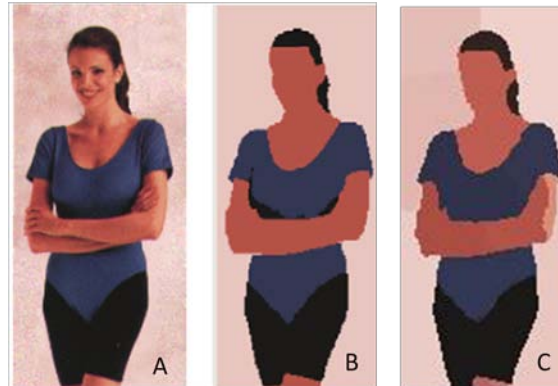


Εικόνα 3.4: Ροή επεξεργασίας της κατάτμησης μετά τη συσσώρευση.

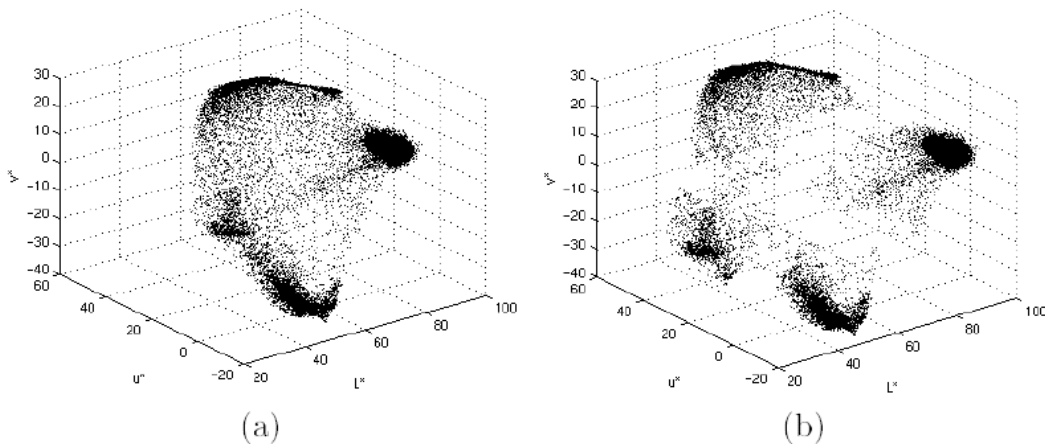
Αξίζει να σημειωθεί ότι οι αντιληπτές χρωματικές διαφορές στο χώρο $L^*U^*V^*$ μετρώνται από Ευκλείδειες αποστάσεις, όπου οι συντεταγμένες του χώρου σχετίζονται άμεσα με τις τιμές RGB μέσω μη γραμμικών μετασχηματισμών όπου το πρότυπο φως της ημέρας D65 χρησιμοποιείται ως αναφορά φωτεινότητας. Η χρωματική πληροφορία κομίζεται από την u^* και v^* , ενώ οι συντεταγμένες φωτισμού L^* μπορούν να θεωρηθούν ως η σχετική φωτεινότητα. Ψυχοφυσικά πειράματα δείχνουν ότι ο χώρος $L^*u^*v^*$ δεν μπορεί να είναι απόλυτα ισοτροπικός [18], ωστόσο, διαπιστώθηκε ικανοποιητικός για την κατανόηση εφαρμογών εικόνας [5]. Είναι προφανές ότι η αναγωγή των αποτελεσμάτων αλλά και συγκρίσεων γίνεται στον RGB χώρο, για λόγους ευκολίας κατανόησης.

Πείραμα 3. Αυτό το πείραμα κατάτμησης χρησιμοποίησε την έγχρωμη εικόνα 3.5α. Ο χρωματικός χώρος (Εικόνα 3.6α) περιείχε 14826 σημεία, και οι τέσσερις ομάδες συσσώρευσης έχουν εξαχθεί χρησιμοποιώντας μια ακτίνα του $h = 10$. Ας σημειώσουμε τα παράτυπα όρια των ομάδων στην Εικόνα 3.6β. Η ποιότητα ομαδοποίησης μπορεί να εκτιμηθεί με την παρατήρηση της κατατετμημένης Εικόνας 3.5β, όπου έχουν χρησιμοποιηθεί χωρικοί περιορισμοί για την άρση των μικρών περιφερειών που περιέχουν λιγότερα από 25 εικονοστοιχεία. Η δυνατότητα αυτή να μην βοηθάει στις μικρές και όχι ιδιαίτερα πολύπλοκες εικόνες, στις οποίες θέλεις να αναδείξεις την

αφαιρετική ικανότητα του αλγορίθμου, ωστόσο στα γεωγραφικά δεδομένα, η κλίμακα αλλά και η ανάλυση των διαθέσιμων δεδομένων επιτάσσουν μια αρκετά υποκειμενική άποψη στην χρήση τέτοιου ορίου. Όντως, στην εικόνα 3.5c βλέπουμε ότι η κατάτμηση είναι αρκετά ικανοποιητική. Οι δύο διαφορές που εντοπίζονται στο μπράτσο της γυναίκας και στο φόντο της εικόνας, ανήκουν στην ίδια κατηγορία, και δεν βασίζονται στον χωρικό προσδιορισμό των 25 εικονοστοιχείων, αλλά στην χρωματική ευαισθησία του αλγόριθμου.



Εικόνα 3.5: Τρίτο πείραμα. (A) Αρχική έγχρωμη εικόνα. (B) Κατατετμημένη εικόνα από το Rutgers University (C) Κατατετμημένη από τον υπό αξιολόγηση αλγόριθμο



Εικόνα 3.6: Τρίτο πείραμα. (α) Αρχικό σύνολο δεδομένων (14826 χρωματικά σημεία) αντιστοιχούν στην εικόνα 3.5 α. (β) Οριοθέτηση ομαδοποίησης (4 ομάδες συσσώρευσης). Η τοποθεσία κάθε ομάδας έχει μετατοπισθεί για να δείξει την οριοθέτηση.

Δοκιμάστηκε η σταθερότητα του αλγορίθμου χρησιμοποιώντας διαφορετικά σύνολα δειγματικών σημείων, με κάθε σύνολο να οδηγεί σε μια ξεχωριστή ψηφίδωση του εισαγόμενου χώρου.

Τέσσερις αξίες της ακτίνας παραθύρου h έχουν θεωρηθεί: 4, 7, 16 και 22. Δέκα μελέτες έχουν διεξαχθεί για

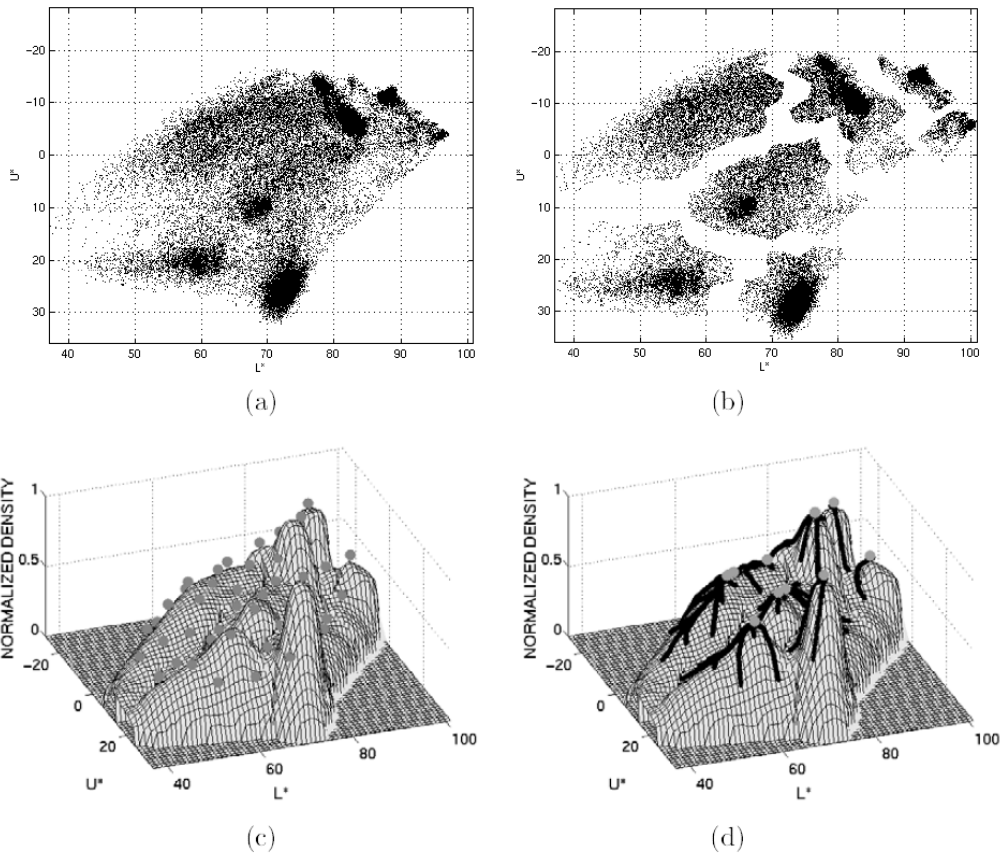
κάθε παραθυρική ακτίνα. Ο αλγόριθμος αποδείχθηκε πολύ σταθερός στην παραγωγή ανάλογων τοποθεσιών και αποτύπωσης των ομάδων συσσώρευσης για κάθε συγκεκριμένη τιμή ακτίνας. Ο πίνακας 3.1 παρουσιάζει τον αριθμό των ανιχνευμένων ομάδων που αντιστοιχούν σε κάθε κατηγορία τιμής της ακτίνας.

Πίνακας 3.1: Αριθμός των ανιχνευόμενων ομάδων έναντι της σφαιρικής ακτίνας για το πείραμα 3.

Detected Clusters	5	4	4	3
Sphere Radius	4	7	16	22

Πείραμα 4. Ένα αρκετά πολύπλοκο στην ανάλυση του, σύνολο δεδομένων φαίνεται στην Εικόνα 3.8 α. Περιέχει 17748 σημεία και αντιπροσωπεύει τα 2 πρώτα στοιχεία του χώρου $L^*u^*v^*$ της έγχρωμης εικόνας 3.9 α. Χρησιμοποιήσαμε μόνο αυτόν τον υπο-χώρο για να μπορούμε να απεικονίσουμε τη συμπεριφορά του αλγορίθμου. Στο αποτέλεσμα του πειράματος, παρατηρείται μεγάλη ποσότητα θορύβου στο φόντο της εικόνας, ασύμμετρες συσσωρεύσεις, περιορισμένες κορυφές και μεγάλες πεδιάδες. Είναι προφανές ότι τα πραγματικά δεδομένα έχουν συχνά τέτοια περίπλοκη δομή.

Χρησιμοποιώντας μία ακτίνα $h = 5$ ο προτεινόμενος αλγόριθμος εντόπισε 7 ομάδες συσσώρευσης (Εικόνα 3.8 b). Τα 47 σημεία δειγματοληψίας εμφανίζονται στην Εικόνα 3.8 c μαζί με την εκτίμηση της πυκνότητας $E_{\text{ρανεχνικών}}$. Η εκτίμηση υπολογίστηκε με μία ανάλυση του ενός εκ των δύο αξόνων και χρησιμοποιώντας την ίδια ακτίνα παραθύρου $h = 5$. Το δειγματικά σημεία συνέκλιναν προς τα υποψήφια κέντρα της συσσώρευσης (Εικόνα 3.8 d) που βρίσκονται σε τοπικά μέγιστες εκτιμήσεις της πυκνότητας. Η εφαρμογή ελέγχου για σημεία κοιλάδας απομάκρυνε όποιες από τις τιμές κεντρικών υποψηφίων της συσσώρευσης που βρίσκονται σε στατιστικά οροπέδια, επιτρέποντας τη σωστότερη οριοθέτηση ομάδων συσσώρευσης.



Εικόνα 3.8: Πέμπτο πείραμα (α) Αρχικό σύνολο δεδομένων. (β) Ανάλυση δεδομένων (σε 7 ομάδες συσσώρευσης). Η τοποθεσία κάθε ομάδας έχει μετατοπισθεί για να δείξει την οριοθέτηση. (γ) Σύνολο δειγμάτων (47 σημεία) και η εκτίμηση πυκνότητας Ερανεχνικών. (δ) Mean Shift τροχιές και υποψήφια κέντρα ομάδων συσσώρευσης.

Η ανάλυση στην Εικόνα 3.8 β δεν έχει φυσική σημασία δεδομένου ότι βασίζεται μόνο σε δύο διαστάσεις. Ωστόσο, όταν η έγχρωμη εικόνα β ~~π~~ρεκειται υπό επεξεργασία στον πλήρη (τριδιάστατο) χρωματικό χώρο, η παραγόμενη κατάτμηση είναι ικανοποιητική (Εικόνα 3.9b)



Εικόνα 3.9: (Α) Αρχική εικόνα. (Β) Κατατεμημένη εικόνα από το Rudgers University (C) Κατατεμημένη από τον υπό αξιολόγηση αλγόριθμο

4 Google Earth Παρουσίαση περιβάλλοντος προγραμματισμού, εφαρμογή κατάμτησης στο Google Earth Εισαγωγή

4.1.1 Η τεχνολογία του Google Earth

Χρησιμοποιώντας το GE οι άνθρωποι έχουν δεδομένη τη δυνατότητα να εστιάζουν, να περιστρέφουν, να δίνουν κλίση και να επιλέγουν την κάτοψη ή το



Εικόνα 4.1 – Η πολυγωνική εικονική πραγματικότητα

προφίλ μιας εικόνας που μπορεί σε λεπτομέρεια να φτάνει και την αυλή του σπιτιού τους. Αλλά η πατέντα του GE δεν είναι στην σύνθεση και αποθήκευση αυτών των εικόνων – είναι στην γρήγορη και αποτελεσματική απόδοση αυτών στον υπολογιστή του κάθε χρήστη. Δεδομένου ότι με μια απλή σύνδεση μόντεμ, ταχύτητας 56k ο χρήστης θα ήθελε περίπου 12.400 χρόνια για να κατεβάσει μία εικόνα ανάλυσης 1 μέτρου, για όλο τον κόσμο, το GE δίνει την αίσθηση ότι αυτή η εικόνα, απλά υπάρχει μπροστά στο χρήστη. Η λύση στο πώς ακριβώς θα μεταφερόταν τόσο γρήγορα, τόση πληροφορία, ήταν ένα από τα μεγαλύτερα εμπόδια που κλήθηκε να ξεπεράσει αυτό το λογισμικό.

Μία από τις λύσεις για να μειωθεί ο χρόνος μεταφοράς της πληροφορίας, είναι με το να αξιοποιηθεί ο δίσκος cache του υπολογιστή. Για λόγους απλοποίησης, ο δίσκος cache μπορεί να περιγραφεί σαν προσωρινός χώρος αποθήκευσης του υπολογιστή για αρχεία που ο υπολογιστής έχει ήδη προσπελάσει. Αν ο υπολογιστής μπορεί να αποθηκεύσει εικόνες που έχουν ήδη χρησιμοποιηθεί από το GE, τότε είναι προφανές ότι δεν χρειάζεται να τις ξανακατεβάσει ο χρήστης όταν θελήσει να τις ξαναδεί. Πάνω σε αυτή τη λογική στηρίζεται αρκετά το GE.

Ακόμα ένα βήμα παραπέρα, το να έρθει όλη η πληροφορία στην οθόνη του χρήστη, έχει άμεση σχέση με το πώς ο υπολογιστής επεξεργάζεται και προβάλλει τους χάρτες. Η εταιρία Google κατέχει πολλά πατενταρισμένα λογισμικά, που υποβοηθούν σε αυτό το GE. Ένα από αυτά που πραγματικά κάνει τη διαφορά είναι το επονομαζόμενο Universal Texture (UT), το οποίο διενεργεί δύο πολύ σημαντικές διαδικασίες στον υπολογιστή: την **mip mapping** και την **clip stacking**

Mipmaps είναι η συλλογές από τις εικόνες bitmap, οι οποίες δουλεύουν μέσα στην υφή της εικόνας, ώστε να δημιουργήσουν την ψευδαίσθηση του βάθους. Δουλεύουν με τη λογική της αντίστροφης πυραμίδας, καθώς η κάθε μια είναι «κολλημένη» πάνω στην άλλη έχοντας διπλάσια διακριτική από την προηγούμενη και μισή από την επόμενη. Στην τελική μορφή της αυτή η πυραμίδα, δίνει μια τρισδιάστατη μορφή σε μια δισδιάστατη εικόνα.

Clip stacks είναι ένα μικρό μέρος των τεράστιων mipmaps τα οποία είναι προσαρτημένα σε ένα συγκεκριμένο μέγιστο μέγεθος. Το πλεονέκτημα της σύνδεσης αυτών με τα τεράστια αρχεία είναι απλό. Το GE στηρίζεται στο γεγονός ότι οι χρήστες, χρειάζεται να βλέπουν ένα μόνο μικρό κομμάτι

του mirmap κάθε στιγμή. Όταν ο χρήστης λέει στο πρόγραμμα που να «πετάξει», το πρόγραμμα χρησιμοποιεί έναν αλγόριθμο για να κατανοήσει ποιο κομμάτι ακριβώς του μεγάλου χάρτη επιθυμεί να δει ο χρήστης, ώστε να ενεργοποιήσει μόνο τα συγκεκριμένα clips και να δημιουργήσει την τελική εικόνα που φαίνεται στην οθόνη.

Τοποθετώντας το όσο πιο απλά γίνεται, το GE έχει δημιουργήσει μια τεράστια, πολλών terabyte, πολύ υψηλής ανάλυσης δορυφορική κάλυψη όλης της γης, και την αποδίδει σε πολύ μικρά κομμάτια, ώστε να είναι «εύπεπτα» από τον μέσο υπολογιστή. Με το ζουμ στην αυλή του σπιτιού του χρήστη, ο υπολογιστής πετάει εκτός όλη την υπόλοιπη πληροφορία που δεν αφορά την αυλή αυτή.

Αυτό ακριβώς ήταν και το πρώτο βήμα στην κατανόηση του τρόπου προσέγγισης στην χρήση πολλών εικόνων του GE με ένα πιο αυτοματοποιημένο τρόπο.

4.1.2 Παραμετροποιώντας το Google Earth

Ένα από τα σημαντικότερα πράγματα που προσφέρει το GE είναι ο βαθμός επέμβασης και παραμετροποίησης που δίνει στον χρήστη, ακόμα και στην απλή – δωρεάν έκδοση του λογισμικού. Με μεγάλη ευκολία μπορεί ακόμα και ένας αρχάριος χρήστης να τοποθετήσει κάποια σημεία ενδιαφέροντος, ή να σχηματίσει κάποια πολύγωνα σε περιοχή που θέλει να ξανα-εντοπίσει αργότερα, ή ακόμα και να δημιουργήσει επικάλυμμα στη δορυφορική εικόνα, μιας φωτογραφίας που έχει επιλέξει. Με την ίδια λογική μπορεί να σχεδιάσει διαδρομές πάνω στο χάρτη, σύνορα, κτλ. Όλα αυτά βασίζονται στην γνωστή πλέον γλώσσα KML η οποία επιτρέπει τη διάδραση της εξωγενούς πληροφορίας, με την πληροφορία που βρίσκεται μέσα στο GE.

Αυτή η γλώσσα ήταν και η τελική σύνδεση του περιβάλλοντος διάδρασης του προγράμματος (Graphic User Interface – GUI) που χρησιμοποιήθηκε στην παρούσα πτυχιακή, για την οριοθέτηση, και στη συνέχεια εξαγωγής, της περιοχής έρευνας

4.2 Το λογισμικό Earthwatch

4.2.1 Introduction

Η ανάπτυξη του συγκεκριμένου λογισμικού, ήταν ταυτόχρονη με την εκπόνηση της παρούσας εργασίας, ενώ μέχρι την ημέρα παράδοσης της τελευταίας βρισκόταν ακόμα σε beta (δοκιμαστικό) στάδιο. Ο υπεύθυνος προγραμματιστής Ian Roessle, συνέβαλε στα προγραμματιστικά θέματα του GE, ενώ σε αντάλλαγμα, το λογισμικό του, απέκτησε καινούριες δυνατότητες, επαληθευτικούς αλγορίθμους, και φυσικά το κομμάτι της κατάτμησης της εικόνας.

Ξεκινώντας το πρώτο προφανώς απαραίτητο πράγμα που πρέπει να προηγηθεί όλων των άλλων για να διενεργηθεί ή όποια έρευνα, είναι να «κομιστεί» η εικόνα της περιοχής ενδιαφέροντος. Το

GE δίνει την περιορισμένη δυνατότητα να εξάγεις μια εικόνα από 1000 ως 4800 pixels (ανάλογα με την έκδοση που έχει στην κατοχή του ο χρήστης) ωστόσο στη διαδικασία αυτή έχει δύο σημαντικά μειονεκτήματα.

- Αποθηκεύεται σε lossy - JPEG μορφή
- Για μια περιοχή με τάξη μεγέθους μιας πόλης σαν το Las Vegas, και μέγεθος εικονοστοιχείου 1 μέτρου, χρειάζονται περίπου 1764 εικόνες οι οποίες θα σωθούν χειρωνακτικά, μία μία για να γίνει η επεξεργασία κτλ.

Μέσα από το Earthwatch, υπάρχει η δυνατότητα να γίνει αυτόματα το μέγιστο δυνατό ζουμ (άρα και να μη χαθεί καμία πληροφορία) και φυσικά να σωθεί σε μορφή bitmap.

Στο επόμενο βήμα, με την χάραξη κάποιων πολυγώνων, μέσα στην υπό μελέτη περιοχή, μπορεί να γίνει η απόδοση περιγραφικών ετικετών σε οποιαδήποτε αντικείμενο ή γεωμορφή πάνω στην εικόνα. Αυτό θα αποτελέσει και την περιοχή εκπαίδευσης του αλγορίθμου κατηγοριοποίησης. Η υπολογιστική όραση και η αυτόματη αναγνώριση εικόνας, είναι ένας ανοιχτός και ταχύτατα αναπτυσσόμενος τομέας έρευνας. Αν και οι δυνατότητες κατηγοριοποίησης δεν αποτελούν αντικείμενο της παρούσας εργασίας, ωστόσο θα γίνει εκτενής αναφορά στους αλγορίθμους (machine learning) οι οποίοι χρησιμοποιούνται[Παράρτημα I].

Το Earth Watch δεν δίνει τη δυνατότητα μιας μεθόδου κατηγοριοποίησης μόνο, παρά καλεί ένα ξεχωριστό πακέτο Machine Learning, το οποίο ονομάζεται WEKA_2

Το στοιχείο/κριτήριο κατηγοριοποίησης του κάθε εικονοστοιχείου αποτελείται από μια ενημέρωση χρώματος, μαζί με μια ποικιλία πληροφορίας διάφορων επιπέδων επηρεασμού (weighted). Παραδείγματος χάρη, η πληροφορία σχετικά με τα γειτονικά εικονοστοιχεία για το πού βρίσκονται σε κοντινή απόσταση, επηρεάζουν σαφώς περισσότερο την κατηγοριοποίηση του συγκεκριμένου pixel, απ' ό,τι αυτά που βρίσκονται πιο μακριά. Επί του παρόντος, η κατηγοριοποίηση αυτή χρησιμοποιείται μόνο σαν μέσο για την διάκριση αρκετά ομοιογενών (χρωματικά) χαρακτηριστικών.

Με αυτό το επίπεδο ανάλυσης θα μπορούσε κάποιος να κάνει μια μελέτη του ποσοστού αλλαγής συγκεκριμένων βιομαζών, όπως παραδείγματι το ποσοστό αποψίλωσης σε ένα τροπικό δάσος. Ενώ ένα μεγάλο μέρος της συζήτησης σχετικά με τη συχνότητα των ενημερώσεων των εικόνων στο Google Earth είναι καθαρά υποθετικό, θα μπορούσε να θεωρηθεί ότι σε διάρκεια 3 ετών γίνεται, σίγουρα, κάποια ανανέωση στη βάση δεδομένων σχεδόν όλου του κόσμου. Σύμφωνα με τις διάφορες παρατηρήσεις οι ανανεώσεις γίνονται στην πλειοψηφία τους τη 15^η ημέρα του κάθε μήνα. Αυτή είναι και η μόνη σχετικά επίσημη ενημέρωση αυτή τη στιγμή στο διαδίκτυο από την GE. Ανεξάρτητα όμως, με το ρυθμό τον οποίο οι εικόνες ανανεώνονται, κάνοντας την σύγκριση των αποτελεσμάτων ενός πειράματος, στην ίδια γεωγραφική περιοχή, με εικόνες που λήφθηκαν σε διαφορετικές ημερομηνίες θα μπορούσε να καθοριστεί το ποσοστό αλλαγής των αναλογιών

(change detection) στις υπό μελέτη βιομάζες – πάντα για τη συγκεκριμένη χρονική περίοδο. Τα αποτελέσματα εξάγονται σε μορφή PDF στο τέλος κάθε ολοκληρωμένου πειράματος που παρουσιάζει γεωγραφική περιοχή πριν και μετά από την ταξινόμηση, δίνοντας ένα κωδικοποιημένο χρώμα για την κάθε κλάση της κατηγοριοποίησης που έλαβε χώρα. Επίσης γίνεται και ο υπολογισμός απόδοσης του ταξινομητή για ευκολότερη αποτίμηση της κατηγοριοποίησης.

4.2.2 Τεχνικό Υπόβαθρο

Στο πειραματικό στάδιο το EarthWatch είχε γραφτεί σε γλώσσα C #, καθώς αυτό εξασφαλίζει συμβατότητα με μια μεγάλη γκάμα εργαλείων για τη σύνδεση του λογισμικού με Win32 COM και GDI δύο τεχνολογίες που αποτελούν τους βασικούς πυλώνες του προγράμματος. Τελικά ωστόσο ο κώδικας μεταφέρθηκε σε Java Second Edition 6,0 που είναι και η γλώσσα που αποτελεί τον πυρήνα της βάσης του κώδικα. Η γλώσσα Java επιλέχθηκε για διάφορους λόγους, με κύρια αιτία τους πολλούς ήδη ανεπτυγμένους αλγορίθμους, ανοιχτού κώδικα, που χρησιμοποιήθηκαν στο σύστημα. Επίσης, το περιβάλλον ανάπτυξης, NetBeans, μια δημιουργία της Sun Microsystems, παρείχε ένα πολύ εύχρηστο εργαλείο για την ανάπτυξη και το σχεδιασμό του Swing JFrame, την ταχεία διαμόρφωση κώδικα, την ολοκλήρωση και την αυτόματη αύξηση και ανάπτυξη ταχύτητας στο σύνολο της εφαρμογής.

Προκειμένου να γίνει η σωστή ενοποίηση της γλώσσας, όντας ανεξάρτητη της πλατφόρμας ανάπτυξης, στο απαιτούμενο περιβάλλον, ένα κομμάτι της βάσης του κώδικα έπρεπε να είναι γραμμένη σε Native Win32 [Παράρτημα I] C και να συνδεθεί με την κύρια προγραμματιστική βάση μέσω του Java Native Interfaces (JNI) [Παράρτημα I]. Ο κώδικας C έχει συνταχθεί σε ένα Dynamic Link Library (DLL) [Παράρτημα I] που φορτώνεται στις classes του κώδικα ως μητρική λειτουργία. Έτσι το λογισμικό έχει την ευελιξία να έχει πρόσβαση σε ειδικά παράθυρα, ρουτίνες, για να διευκολυνθεί η ενσωμάτωση του, με το Google Earth. Μία από τις πιο χρήσιμες βιβλιοθήκες Windows είναι Shell32 [Παράρτημα I] η οποία δίνει τη δυνατότητα της πρόσβασης σε φακέλους του συστήματος και συγκεκριμένα στο φάκελο Application Data. Άλλη βιβλιοθήκη είναι η gdi32, ή Graphics Device Interface [Παράρτημα I] η οποία επιτρέπει την άμεση διεπαφή του χρήστη με τις ιδιότητες ενός περιβάλλοντος διάδρασης. Τέλος, User32 χρησιμοποιήθηκε προκειμένου να διευκολυνθεί η προσομοίωση των πλήκτρων του πληκτρολογίου. Επιπροσθέτως έγινε και η χρήση του COM4J, ενός βοηθητικού υποπρογράμματος, το οποίο επέτρεψε τη σύνδεση του COM interface του Google Earth στον Java κώδικα. Στην πραγματικότητα δημιουργεί αυτόματα ένα σύνολο από κλάσεις Java, ανάλογα με την πληροφορία που παίρνει από το GE.

Όσον αφορά τα πειράματα τώρα, με τη δημιουργία τους, αποθηκεύονται σε απλό αρχείο στο φάκελο εγγράφων, ο οποίος περιέχει την εικόνα της περιοχής μελέτης, τα πολύγωνα (KML) εκπαίδευσης του κατηγοριοποιητή, την κατάσταση στην οποία έχει μείνει το πείραμα, τα

αποτελέσματα με τις αντίστοιχες των καινούριων πολύγωνων και τις σχετικές εκτιμήσεις / αναλογίες της κατηγοριοποίησης.

Το αρχείο KML κωδικοποιείται σε ένα XML όπως ένα Document Type Definition (DTD). Αυτά τα αρχεία KML περιέχουν σημεία θέσης, πολύγωνα, και επιστρώσεις, και μεταξύ άλλων, ό,τι αποτελεί "χαρακτηριστική ιδιότητα" μέσα στο περιβάλλον του Google Earth. Ένα βοηθητικό πρόγραμμα γραμμένο σε Java από τον Keith Power [gekmlib – Παράρτημα I] προβλέπει μια σχετικά εύκολη ανάλυση αυτών των αντικειμένων/δεδομένων – κάτι που θα επέτρεπε και την περεταίρω αξιοποίησή τους. Η τρέχουσα έκδοση ωστόσο - κατά την διάρκεια της συγγραφής αυτής της πτυχιακής- δεν υποστηρίζεται από αυτή τη λειτουργία, χωρίς να αποκλείει όμως το ενδεχόμενο κάποιο από τα updates του κώδικα να το συμπεριλάβει σαν δυνατότητα.

Όσον αφορά την ταξινόμηση και αναγνώριση της εικόνας, είναι προφανές ότι δεν υπάρχει κάποιος ιδεατός αλγόριθμος αναγνώρισης προτύπων ο οποίος επιτυγχάνει σε όλα τα σενάρια που χρησιμοποιείται. Η έρευνα για την εύρεση του βέλτιστου αλγορίθμου ταξινόμησης της εικόνας, δεν αποτελεί στόχο της παρούσας εργασίας, έτσι χρησιμοποιήθηκε το προαναφερθέν πακέτο εργαλείων Weka που περιέχει μια συλλογή αλγορίθμων ταξινόμησης, γραμμένο σε Java. Με τον τρόπο αυτό θα μπορούσε να προστεθεί στο πρόγραμμα ο οποιοσδήποτε αλγόριθμος, για να υπάρχει η δυνατότητα υποστήριξης περισσότερων μεθόδων ταξινόμησης.

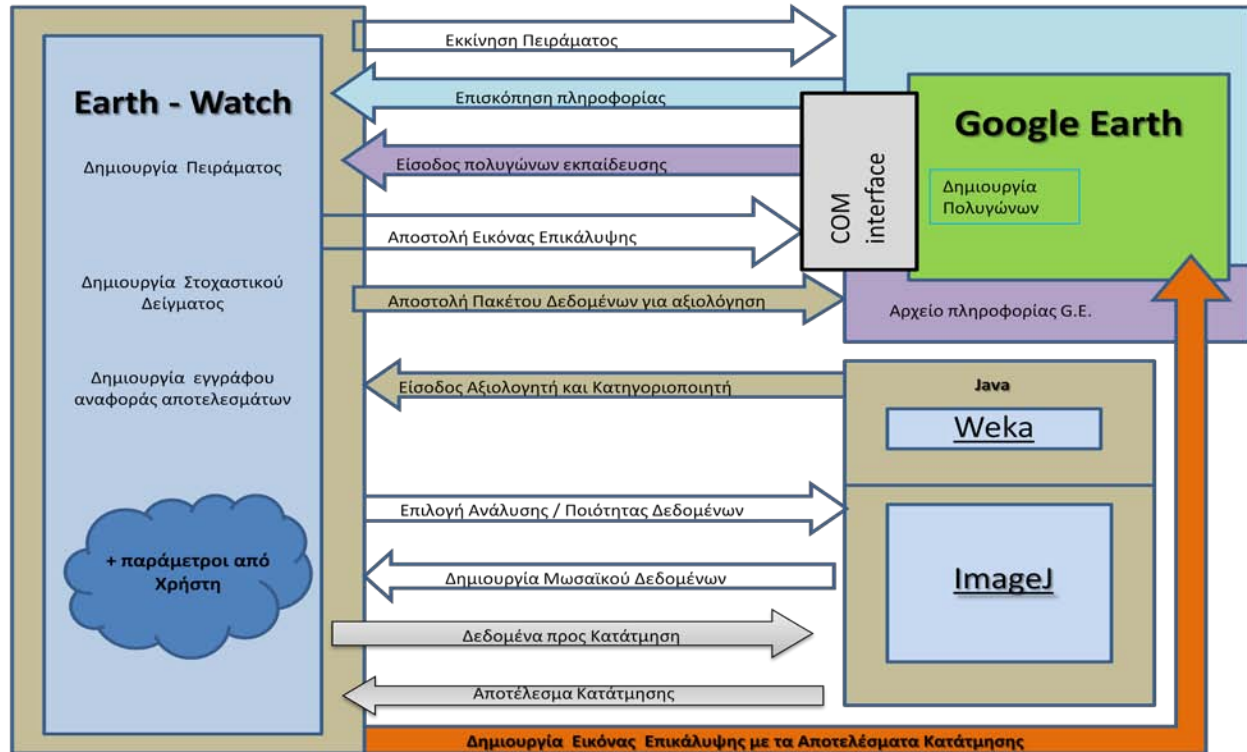
Η τελευταία αλλά σημαντική τεχνική λεπτομέρεια, είναι η απόδοση συντεταγμένων στην εξαγόμενη εικόνα. Αρχικά η Javaloution και οι βιβλιοθήκες της JScience διερευνήθηκαν, καθώς είναι πακέτα με μια μεγάλη συλλογή εργαλείων μετάφρασης μεταξύ των διαφόρων συστημάτων συντεταγμένων. Το μεγάλο πρόβλημα ωστόσο ήταν ότι δεν παρέχει καμία μέθοδο για τον υπολογισμό των αποστάσεων μεταξύ 2 σημείων, οπότε δεν ήταν από πολύ χρήσιμο. Βασιζόμενος στην εξίσωση haversine έγινε εφικτή αυτή η μέτρηση.

Το σύστημα συντεταγμένων που χρησιμοποιήθηκε, ήταν το WGS84, με την χρήση του σφαιρικού αντί του ελλειψοειδούς μοντέλου, για μείωση της υπολογιστικής ανάγκης, καθώς και τον μερικό περιορισμό σφάλματος, τάξης του 10%, λόγω της επικάλυψης μεταξύ των εικόνων που απαρτίζουν την περιοχή έρευνας.

4.2.3 Συνοπτικό Διάγραμμα Ροής Δεδομένων του EarthWatch

Το EarthWatch, παρέχει την δυνατότητα να διεξαχθεί ένα πείραμα στο περιβάλλον, και χρησιμοποιούντες την πληροφορία του Google Earth. Περιληπτικά, οι κύριες συνιστώσες του λογισμικού, παρατίθενται εδώ, σε ένα εικονικό διάγραμμα ροής, που δίνει μια πιο ακριβή αίσθηση της διεξαγωγής ενός πειράματος.

Συνοπτικό σχήμα λειτουργίας του EarthWatch



Εικόνα 4.2. Συνοπτικό σχήμα λειτουργίας του EarthWatch

4.2.3.1 Δημιουργία Πειράματος

Η ενότητα: Δημιουργία; Πειράματος, δημιουργεί το αρχικό αρχείο δεδομένων ενός πειράματος. Οι απαιτούμενες πληροφορίες που απαιτούνται από το χρήστη περιλαμβάνουν το όνομα του πειράματος, τον κατάλογο των κατηγοριών/κλάσσεων για την ταξινόμηση στο πλαίσιο του πειράματος και την περιοχή (study area) στην οποία το πείραμα πρόκειται να πραγματοποιηθεί.

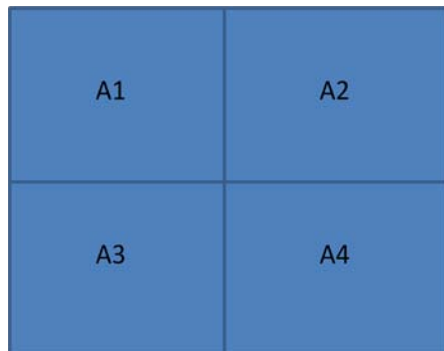
4.2.3.2 Εκκίνηση Πειράματος

Μετά τη δημιουργία του πειράματος, ο χρήστης έχει τη δυνατότητα να το ανοίξει/εκκινήσει. Άμεσα υποδεικνύεται από το λογισμικό η περιοχή μελέτης που είχε οριστεί προηγουμένως, με ένα ημιδιαφανές πολυγωνικό στοιχείο (KML), το οποίο θα αποτελεί και τη βάση για τη συνέχεια της διεξαγωγής του πειράματος στο Google Earth.

4.2.3.3 Επισκόπηση πληροφορίας

Το επόμενο βήμα κατά τη διεξαγωγή του πειράματος είναι να διεξάγει μια έρευνα πάνω από την περιοχή πείραμα. Για την πλήρη επισκόπηση της πληροφορίας, ο χρήστης θέτει το επίπεδο ποιότητας στο οποίο θέλει να διεξάγει την έρευνα, με τον ορισμό της εστίασης (zoom) στην περιοχή. Δημιουργία/είσοδος πολυγώνων εκπαίδευσης.

Στη συνέχεια, γίνεται η χρήση των stitching αλγορίθμων στην βιβλιοθήκη ImageJ, για να ξεκινήσει η δημιουργία του μωσαϊκού των εικόνων. Συγκεκριμένα, δεδομένης της ανάλυσης που επιλέγει ο χρήστης, το πρόγραμμα «τεμαχίζει» την περιοχή μελέτης σε ορθοκανονικό κανάβο, και αρχίζει την αντιστοίχιση εικόνων σε ζευγάρια.



Εικόνα 4.3 – Παράδειγμα κανάβου εικόνας εισαγωγής

Στην περίπτωση που η εφαρμογή απαιτούσε ένα 2X2 κανάβο, όπως στο παραπάνω σχήμα, τότε ο αλγόριθμος, θα ξεκινούσε την διαδικασία stitching πρώτα με το ζευγάρι A1 + A3, στη συνέχεια A2+A4 και τέλος θα ένωνε τα δύο αποτελέσματα. Στο παράδειγμα που ακολουθεί φαίνεται καθαρά η λειτουργία του αλγορίθμου.



Εικόνα 4.4 – Εφαρμογή του αλγορίθμου stitching

Ένα πολύ χρήσιμο progress bar, μας ενημερώνει για το ποσοστό του μωσαϊκού που έχει ολοκληρωθεί. Δυστυχώς, ανάλογα με την τεχνολογία του υπολογιστή στον οποίο γίνεται το πείραμα, ο χρόνος της υπορουτίνας μπορεί να διαφοροποιηθεί σημαντικά.

Στη συνέχεια, ο χρήστης επιλέγει αν θα κάνει κατηγοριοποίηση ή κατάτμηση στην εικόνα. Στην περίπτωση που θέλει να κάνει classification, τότε θα ορίσει τις περιοχές μέσα στην περιοχή μελέτη, οι οποίες θα χρησιμοποιηθούν σαν δείγματα για την επακόλουθη κατηγοριοποίηση. Ο ορισμός των περιοχών, γίνεται επίσης με πολυγωνικές οντότητες (KML) πάνω στο GE. Παρακάτω, με την υπορουτίνα Get Polygons, το λογισμικό προσομοιώνει συγκεκριμένο συνδυασμό πλήκτρων, τα οποία εξασφαλίζουν την αποθήκευση των πιο ενημερωμένων πολυγωνικών οντοτήτων στην περιοχή μελέτης. Αυτές οι οντότητες εμφανίζονται αυτόματα, σε ένα ημιδιαφανές πολύγωνο πάνω στο GE, για επαλήθευση από το χρήστη.

4.2.3.4 Δημιουργία Στοχαστικού Δείγματος

Αυτή η ενότητα είναι διαθέσιμη, για ένα πείραμα, το οποίο διαθέτει ήδη τα πολύγωνα-δείγματα, για την εκπαίδευση του ταξινομητή, και φυσικά έχει διεκπεραιωθεί και η ανάλυση της περιοχής μελέτης. Ο χρήστης επιλέγει τον αριθμό των δειγμάτων που θα δημιουργηθούν μέσα στο κάθε πολύγωνο. Βασιζόμενο πάνω σε αυτά τα σημεία, το λογισμικό επιλέγει τυχαία τα σημεία που ορίστηκαν, εντός του πολυγώνου εκπαίδευσης. Βασιζόμενο σε αυτά τα σημεία «χτίζει» στη συνέχεια μια διανυσματική οντότητα, η οποία θα είναι ουσιαστικά και το κριτήριο της εκάστοτε κλάσης.

4.2.3.5 Αποστολή Πακέτου Δεδομένων για αξιολόγηση

Με τη συγκεκριμένη υπορουτίνα ξεκινά το πακέτο εργαλείων Weka_2, το οποίο εφαρμόζει μια «10 cross-folded Evaluation»¹. Η συγκεκριμένη αξιολόγηση, γίνεται χρησιμοποιώντας ένα Support Vector Machine αλγόριθμο. Όταν ο ταξινομητής, ολοκληρώσει την ταξινόμηση, σώζει το αποτέλεσμα σε ένα πίνακα σύγχυσης –confusion matrix - , αναπαριστώντας έτσι την ακρίβεια που πέτυχε ο αλγόριθμος, στην κατηγοριοποίηση του κάθε εικονοστοιχείου.

4.2.3.6 Δημιουργία Εγγράφου Αναφοράς

Για την ολοκλήρωση ενός πειράματος στο EarthWatch, δημιουργείται ένα φορητό αρχείο εγγράφου, το οποίο περιέχει μια εικόνα της περιοχής πριν και μια μετά την ταξινόμηση. Συμπεριλαμβάνει επίσης και τον πίνακα σύγχυσης, για την εξαγωγή χρήσιμων στατιστικών στοιχείων.

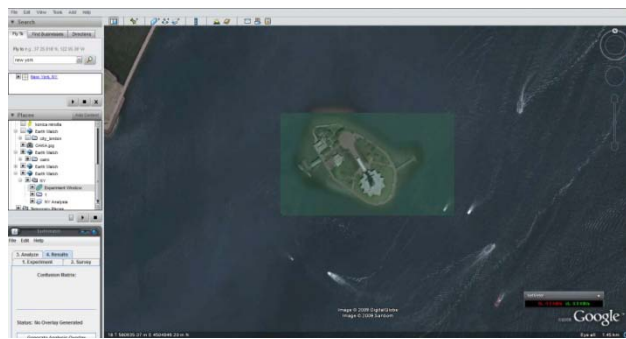
¹ Τοποθέτηση του Link για περισσότερες λεπτομέρειες πάνω στη μέθοδο αυτή.

4.2.3.7 Παράμετροι Κατάτμησης Εικόνας

Με την υπορουτίνα shift, εκ του Mean-Shift, ο χρήστης μπορεί να ανοίξει το διάλογο κατάτμησης που έχει εγκατασταθεί στο πρόγραμμα. Στα αντίστοιχα πεδία θα τοποθετήσει τις παραμέτρους κατάτμησης, Spatial Radius και Color Distance, λαμβάνοντας υπόψη ότι ο πρώτος οφείλει να είναι ακέραιος αριθμός ενώ ο δεύτερος πραγματικός.

4.2.3.8 Δημιουργία Εικόνας Επικάλυψης

Όταν ολοκληρωθεί η διαδικασία, το αποτέλεσμα αυτής έχει αποθηκευτεί σαν αρχείο εικόνας. Επιλέγοντας την δημιουργία εικόνας επικάλυψης, το λογισμικό μας επιτρέπει να ανοίξουμε το αποτέλεσμα απευθείας πάνω στο Google Earth, έχοντας ένα ποσοστό διαφάνειας, για άμεση σύγκριση.



Εικόνα 4.5 – Παράδειγμα δημιουργίας εικόνας επικάλυψης στο περιβάλλον του G.E.



Εικόνα 4.6 –Αριστερά: Αρχικά Δεδομένα, Δεξιά: Αποτέλεσμα Κατάτμησης

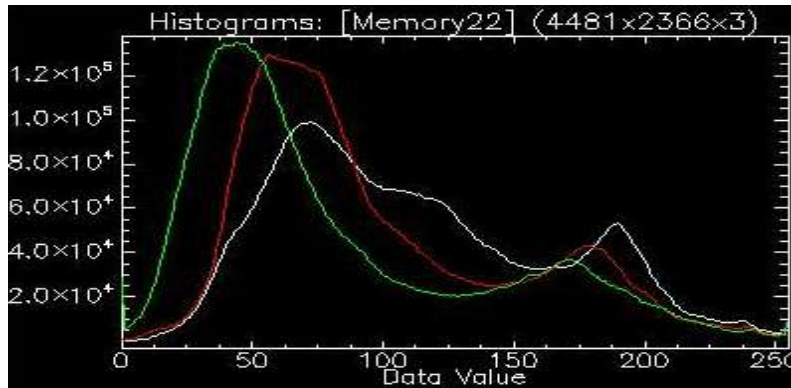
5 Μελέτες Περιπτώσεων Εφαρμογής του αλγορίθμου

5.1 Αεροδρόμιο Ελ. Βενιζέλος, Αθήνα

Πρώτη εφαρμογή του αλγορίθμου, γίνεται στο αεροδρόμιο Ελευθέριος Βενιζέλος. Η αντίθεση χρήσεων γης εδώ είναι μεγάλη, μιας και γύρω από το δομημένο συγκρότημα του αεροδρομίου, υπάρχουν μόνο γεωτεμάχια και το οδικό δίκτυο. Στην παρούσα φάση η κατάτμηση της εικόνας στοχεύει σε μια πρώτη αξιολόγηση του αλγορίθμου σε περιπτώσεις με μεγάλη ποικιλία χρήσεων γης. Σαν τελικό προϊόν, αναμένεται κάποιου είδους συνδυαστική πληροφορία, η οποία με τη σειρά της θα μπορεί να χρησιμοποιηθεί σαν εισροή δεδομένων σε ένα μελλοντικό έμπειρο σύστημα. Ακολουθεί η αρχική [5.1] και η κατατετημημένη [5.3] εικόνα, με τα ιστογράμματα [5.2, 5.4] αυτών.



Εικόνα 5.1– Αρχικά δεδομένα



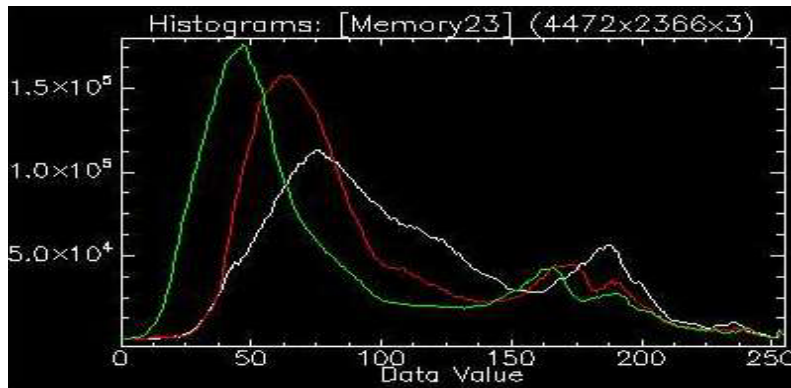
Εικόνα 5.2 . Ιστόγραμμα Χρωματικών τιμών των αρχικών δεδομένων



Εικόνα 5.3 . Κατατετημημένη Εικόνα (ακολουθούν μεγεθύνσεις)

Όπως παρατηρήθηκε από το αποτέλεσμα της κατάτμησης, η χρήση του αλγορίθμου, ναί μεν αφαιρεί το μεγαλύτερο μέρος του θορύβου από την εικόνα, ωστόσο αν κοιτάξουμε σε μεγαλύτερη μεγέθυνση την εικόνα, (παρατίθενται 3 διαφορετικές εκδοχές / επίπεδα κατάτμησης) τότε βλέπουμε ότι η εξαγόμενη πληροφορία, έχει περιορισμένη χρησιμότητα. Για παράδειγμα είναι δυνατόν να γίνει διακριτό ένα αεροπλάνο [5.5.A] (ακόμα κι αν η υφή του αεροδρομίου είναι παρόμοια με αυτή του αεροπλάνου) ένα κτιριακό συγκρότημα [5.5.B], καθώς επίσης και χρήσιμη πληροφορία για μια την υποστήριξη της αποτύπωσης του οδικού δικτύου. Τα ιστογράμματα, συνάδουν με τις παρατηρήσεις, καθώς υπάρχει μια αύξηση στη δημιουργία κορυφών και πεδιάδων, σε βάρος των οροπεδίων που παρατηρούνται στο ιστόγραμμα της αρχικής εικόνας. Ενωιολογικά αυτό σημαίνει ότι ο αλγόριθμος είχε σαν εισερχόμενο δεδομένο μια εικόνα με μια

σχετικά ομαλοποιημένη διασπορά στις τιμές των εικονοστοιχείων της [5.3] και σαν εξαγόμενο δεδομένο [5.4] η εικόνα έχει πλέον κάνει τις απαραίτητες συσσωρεύσεις στα σημεία που η κάθε καμπύλη δημιουργεί κορυφή.



Εικόνα 5.4. Ιστογράμμα Χρωματικών τιμών της κατατετημένης εικόνας



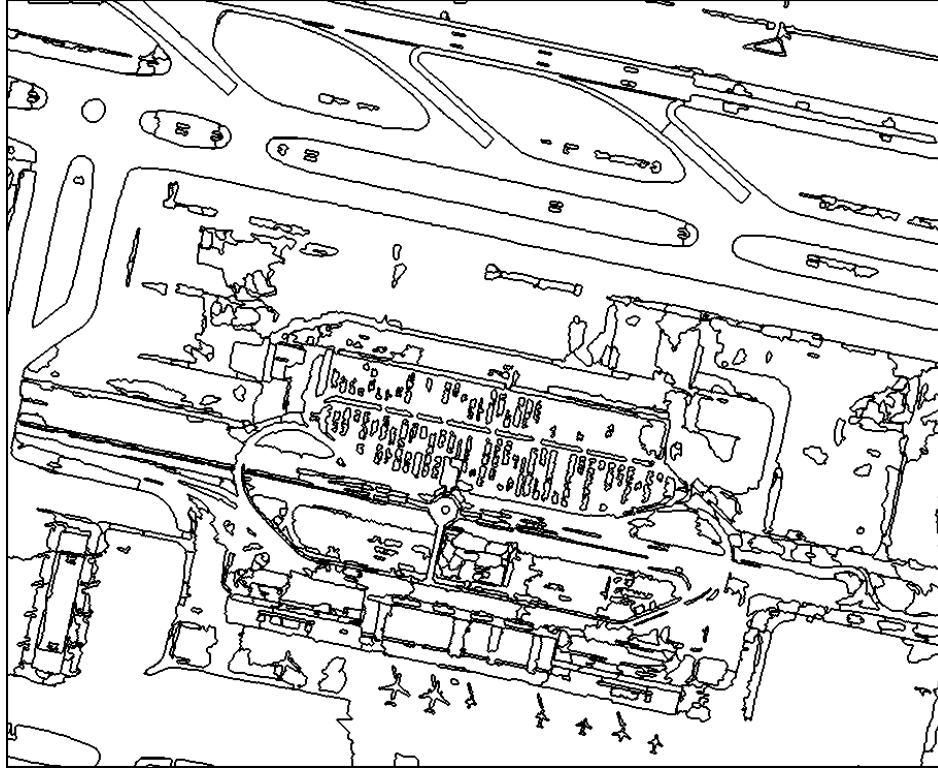
Εικόνα 5.5 - Spatial Radius:10, Color Distance:25



Εικόνα 5.6- Spatial Radius:20, Color Distance:25



Εικόνα 5.7- Spatial Radius:30, Color Distance:20



Εικόνα 5.8- Spatial Radius:30, Color Distance:20, πολύγωνα που δημιουργηθήκαν κατά την κατάτμηση – Όπως είναι ευδιάκριτο, το κεντρικό οδικό δίκτυο, τα μεγάλα κτίρια και αεροπλάνα, κατατμήθηκαν ικανοποιητικά -

Συμπερασματικά, να τονίσουμε την ικανοποιητική απόδοση του αλγορίθμου να αντιμετωπίσει κατασκευές σε εκτός σχεδίου εκτάσεις, όπως τοπικό και εθνικό οδικό δίκτυο, μεγάλες κτιριακές εγκαταστάσεις κτλ.[5.8]. Οι ασάφειες στην κατάτμηση παρουσιάζονται για στοιχεία όπως αυτοκίνητα, ή μη ασφαλτοστρωμένους δρόμους, όπου τα συγκεκριμένα εικονοστοιχεία, σε πολλές περιπτώσεις, κατανέμονται σε λάθος cluster.

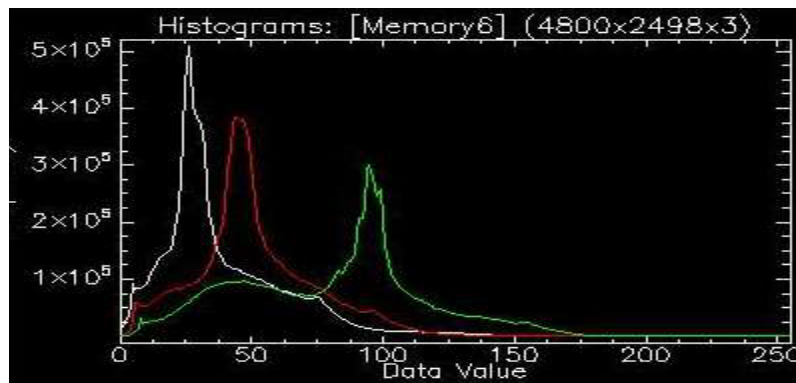
Έστω και με αυτούς τους περιορισμούς ωστόσο, στην πρώτη εφαρμογή του αλγορίθμου, φαίνεται ότι η κατάτμηση μέσα από το συγκεκριμένο λογισμικό, μπορεί να υποβοηθήσει μια περεταίρω κατηγοριοποίηση (ελεγχόμενη ή μη) των εικονοστοιχείων της εικόνας από το χρήστη.

5.2 Big Island, Hawaii

Στη συγκεκριμένη εφαρμογή, γίνεται μια απλοποιημένη προσέγγιση στην χρήση του αλγορίθμου για εξαγωγή γεω-φυσικής πληροφορίας. Η περίπτωση του Big Island της Χαβάης, αναδεικνύει μια τέτοια προσπάθεια, δεδομένου του έντονου ανάγλυφου το οποίο την χαρακτηρίζει. Στα δεδομένα που ληφθήκαν από το Google Earth, υπάρχει και κάποιο μικρό ποσοστό νεφοκάλυψης, το οποίο ο αλγόριθμος θα πρέπει - υπό συνθήκες - να το υποδείξει.



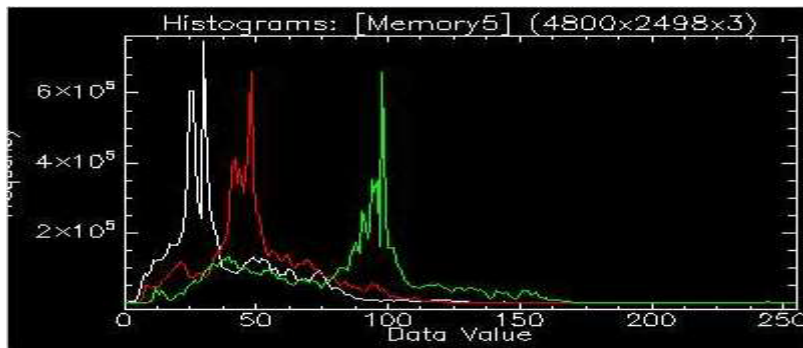
Εικόνα 5.9– Αρχικά δεδομένα



Εικόνα 5.10 Ιστόγραμμα Αρχικών Δεδομένων

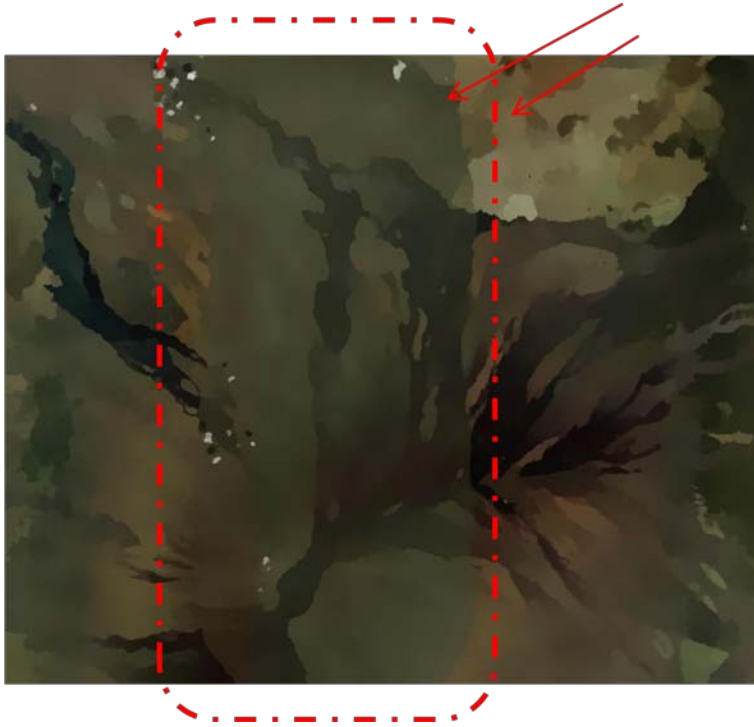


Εικόνα 5.11 Κατάτμηση εικόνας, με συντελεστές Spatial Radius 30, Color Distance 20



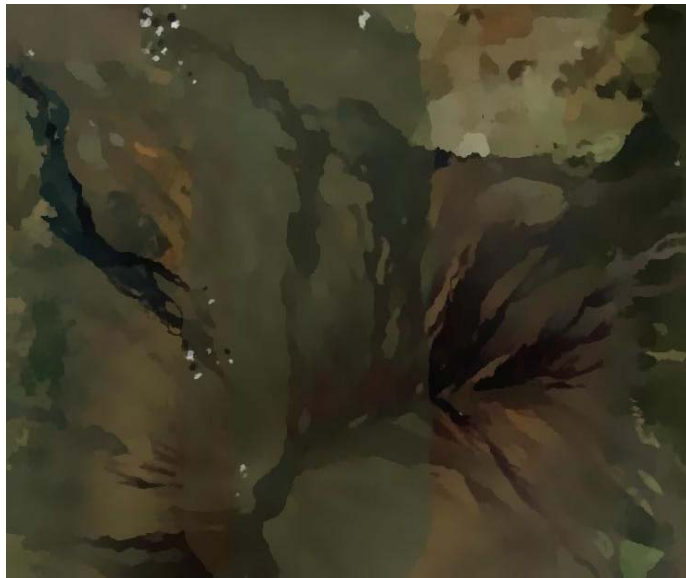
Εικόνα 5.12

Όπως γίνεται αντιληπτό από το ιστόγραμμα των αρχικών δεδομένων, υπάρχει μια κεντρική κορυφή στο κάθε χρώμα, και στη συνέχεια, επιβάλλεται μια ομαλή χρωματική συμπεριφορά στα εικονοστοιχεία. Σε αντίθεση με αυτό, το ιστόγραμμα της κατατετμημένης εικόνας, εμφανίζει αρκετές μεγάλες αλλά μικρότερες κορυφές, υποδεικνύοντας έτσι τις συσσωρεύσεις που δημιουργηθήκαν. Η θάλασσα στο συγκεκριμένο δείγμα, αλλάζει τελείως τη στατιστική της εικόνας, και ενώ έγινε μια προσπάθεια ομογενοποίησης αυτής (στην πραγματικότητα να ομαδοποιηθούν όλα τα εικονοστοιχεία που ανήκουν στην θάλασσα σε ένα cluster) δυστυχώς οι συντελεστές που το κατάφεραν αυτό, ήταν καταστροφικοί και για την υπόλοιπη εικόνα. Δεδομένου ότι η προσπάθεια αυτή στοχεύει στην ανάδειξη χερσαίες γεωμορφές, οι συντελεστές ανάλυσης, τροποποιήθηκαν ανάλογα. Η σκέψη της απομάκρυνσης της θάλασσας από την περιοχή μελέτης, δεν αποτελεί λύση, μιας και σκοπός του πειράματος είναι να αξιολογηθεί ο αλγόριθμος σε ρεαλιστικές συνθήκες έρευνας.



Εικόνα 5.13- Spatial Radius:11, Color Distance:8

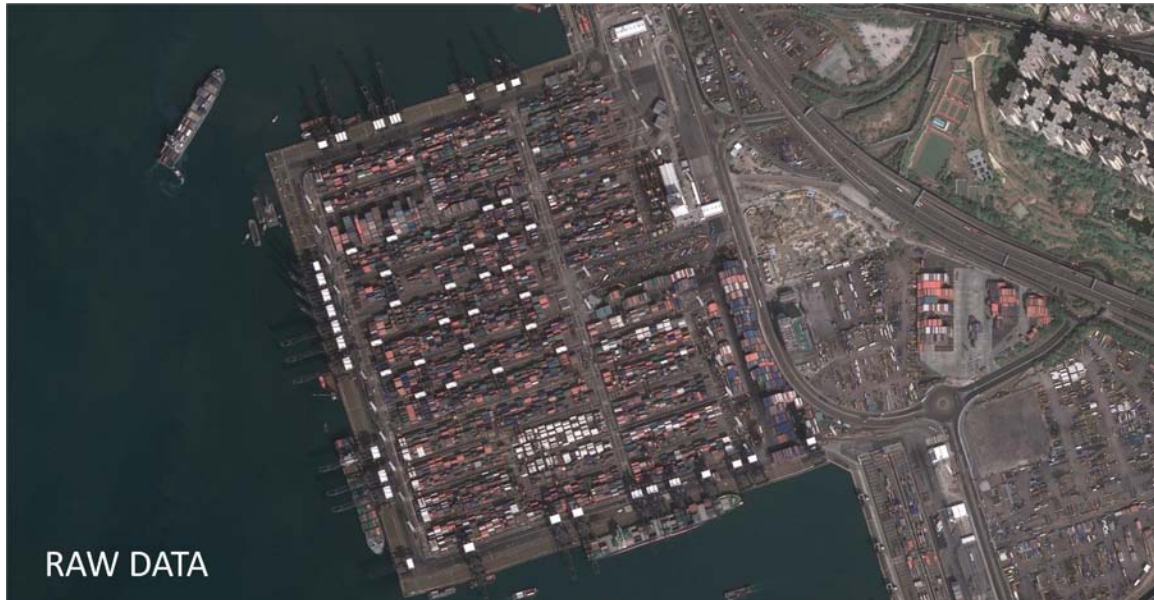
Τα δεδομένα του GE στη συγκεκριμένη περιοχή, προφανώς επειδή έχουν αποτυπωθεί διαφορετική ημερομηνία (μεγάλη ποικιλία στην εδαφοκάλυψη, ηλιοφάνεια, γωνία λήψης δεδομένων, κτλ), έχουν εμφανείς χρωματικές διαφορές [5.13], επηρεάζοντας έτσι την απόδοση του αλγόριθμου.



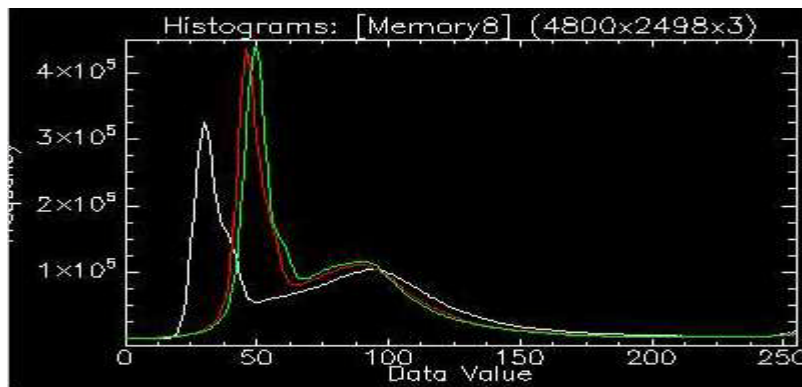
Εικόνα 5.14- Spatial Radius:13, Color Distance:9

5.3 Victoria Harbour, Hong Kong

Μεταφερόμαστε στο λιμάνι Victoria του Χονγκ Κονγκ. Σε μια από της πτέρυγες ενός από τα πιο πολυάσχολα λιμάνια στον κόσμο, (80 – 90% των εξαγωγών της Κίνας, γίνεται από εδώ) θα γίνει η προσπάθεια ανάδειξης των containers που μπαίνουν πάνω στα καράβια. Εδώ το αυστηρά ορθογώνιο σχήμα βοηθάει την κατάτμηση, ωστόσο η ποικιλία χρώματος, δεν φαίνεται να είναι επαρκής.

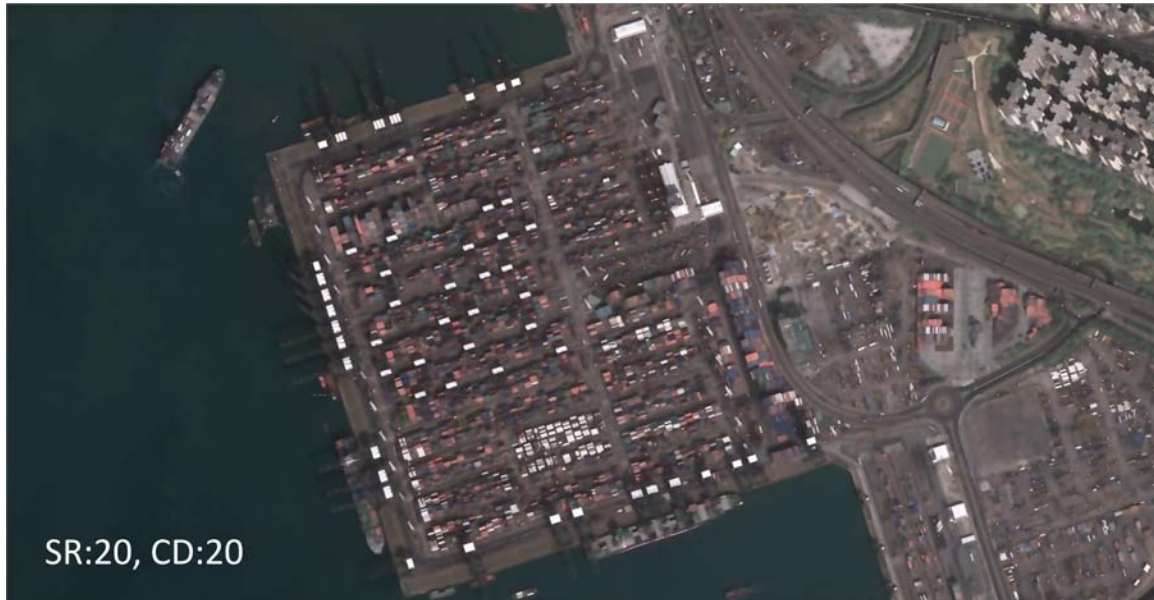


Εικόνα 5.15 – Αρχικά δεδομένα

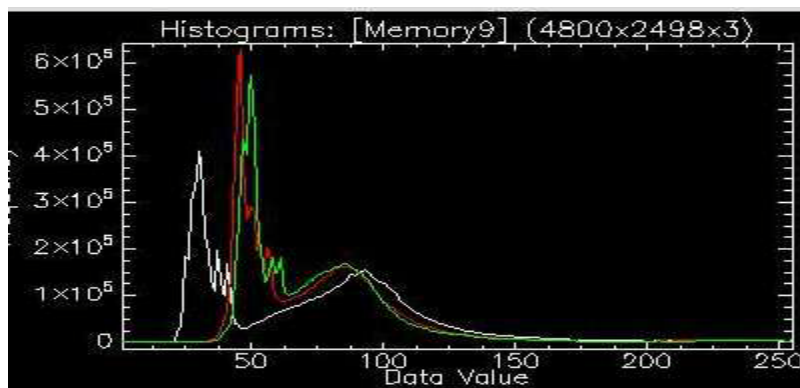


Εικόνα 5.16 – Ιστόγραμμα Αρχικών Δεδομένων

Ακριβώς όπως ήταν αναμενόμενο, πέρα από δύο κορυφές ανά χρώμα, το ιστόγραμμα δεν υποδεικνύει περαιτέρω χρωματικές τιμές που θα μπορούσαν να αποτελέσουν κέντρα συσσωρεύσεων.

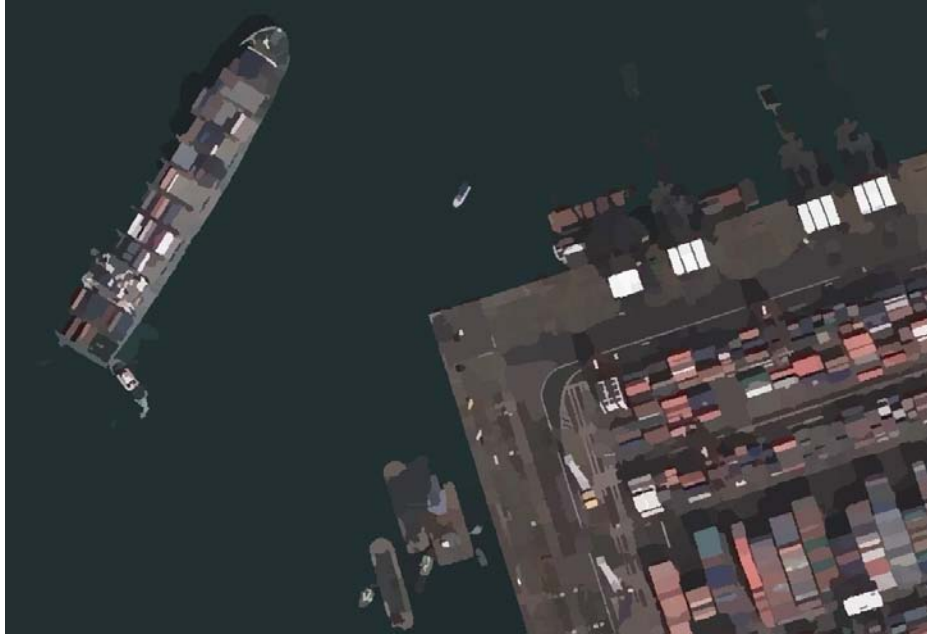


Εικόνα 5.17 – Κατάτμηση εικόνας, με συντελεστές Spatial Radius:20 & Color Distance:20



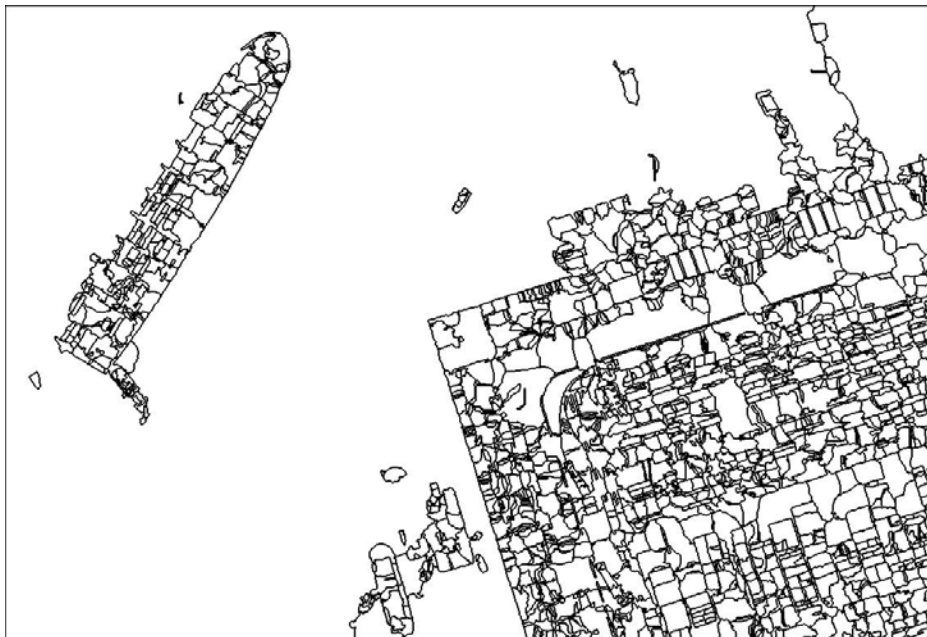
Εικόνα 5.18 – Ιστόγραμμα κατατετημένης εικόνας, με συντελεστές Spatial Radius:20 & Color Distance:20

Μετά από αρκετές προσπάθειες κατάτμησης, το αποτέλεσμα εδώ, αν και στην συνολική εικόνα δεν είναι απολύτως ορατό, δείχνει να παίρνει μορφή. Η αξιολόγηση είναι πολύ καλύτερη σε μεγέθυνση.

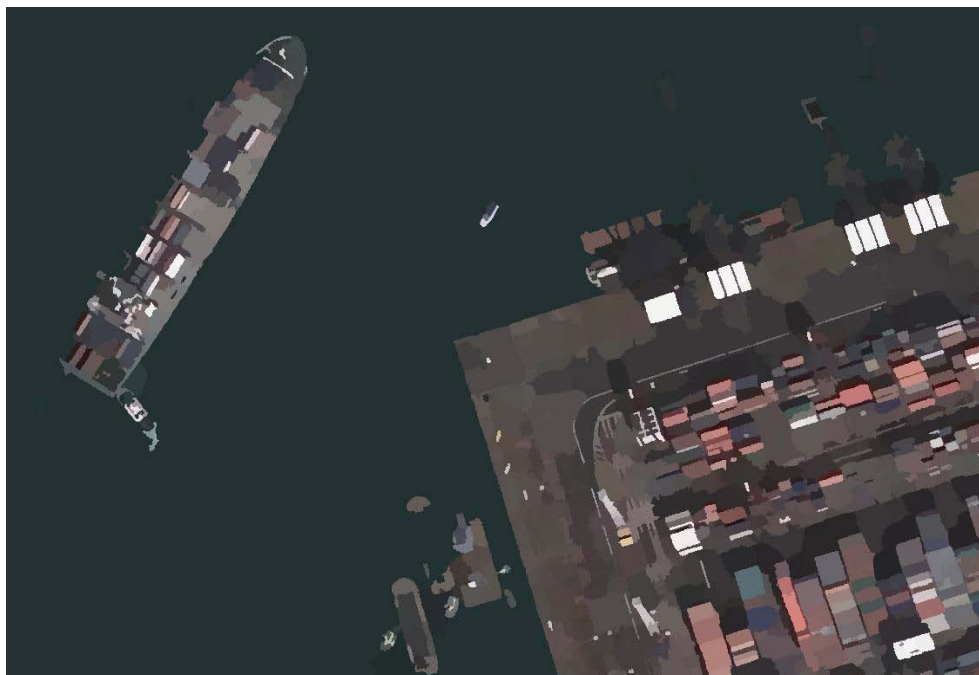


Εικόνα 5.19- Κατάτμηση εικόνας, με συντελεστές *Spatial Radius:11, Color Distance:8*

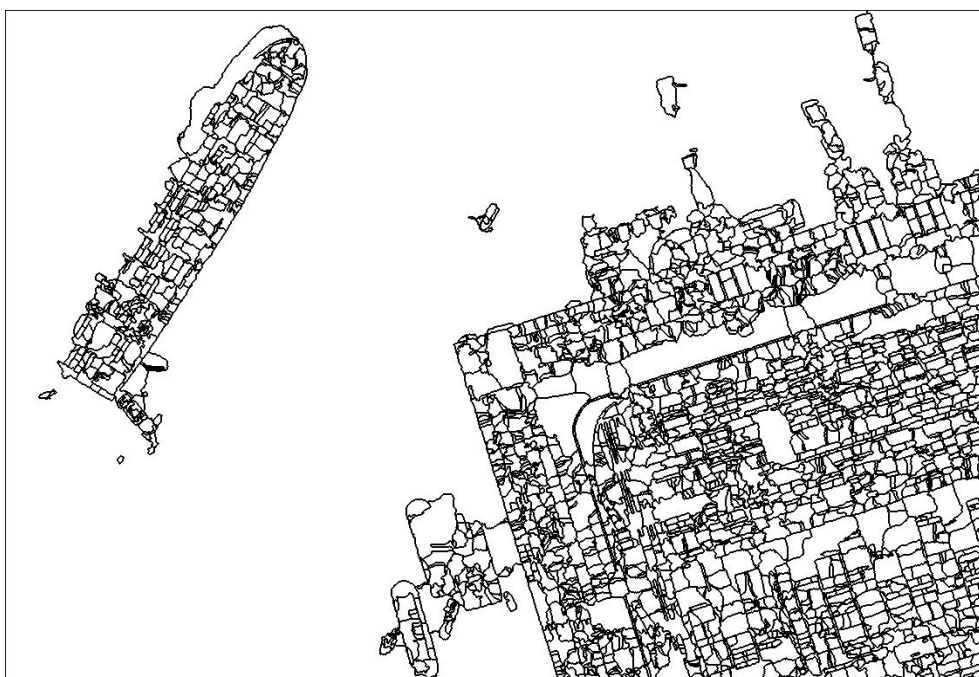
Η θάλασσα πλέον δεν αποτελεί πρόβλημα, έχει κατηγοριοποιηθεί σε ένα cluster, ενώ τα containers, δείχνουν ικανοποιητική διαφοροποίηση, χωρίς ωστόσο να υπάρξει υπερβολική κατάτμηση. Μερική χρωματική σύγχυση εντοπίζεται πάνω στο πλοίο, όπου κάποια από τα κοντέινερς, χάνονται στο χρωματικό υπόβαθρο του καταστρώματος. Ακολουθεί και μια εικόνα με τα όρια των πολυγώνων που δημιουργήθηκαν.



Εικόνα 5.20 Όρια πολυγώνων, *Spatial Radius:11, Color Distance:8*



Εικόνα 5.21- Κατάτμηση εικόνας, με συντελεστές Spatial Radius:9, Color Distance:6



Εικόνα 5.22 Όρια πολυγώνων, κατάτμησης με συντελεστές Spatial Radius:9, Color Distance:6

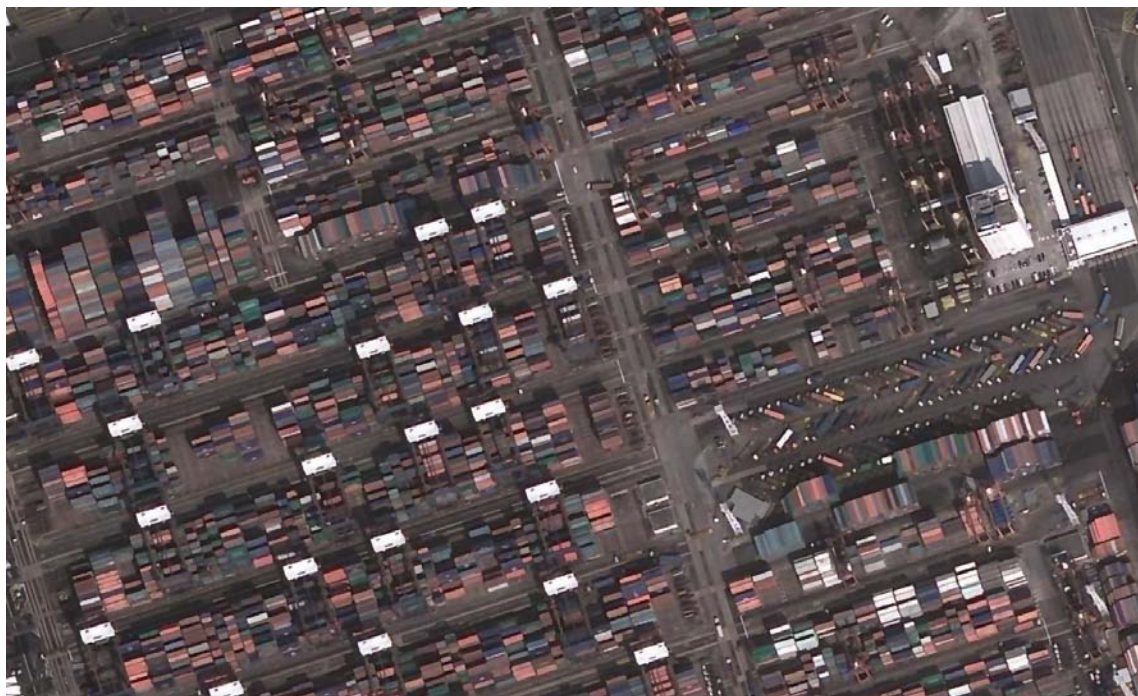
Στις παρακάτω εικόνες, φαίνεται η συμπεριφορά του αλγορίθμου, σε πιο απαιτητικές χρωματικά περιπτώσεις.



Εικόνα 5.23- Κατάτμηση εικόνας, με συντελεστές Spatial Radius:9, Color Distance:6



Εικόνα 5.24 Όρια πολυγώνων, κατάτμησης, με συντελεστές Spatial Radius:9, Color Distance:6



Εικόνα 5.25- Κατάτμηση εικόνας, με συντελεστές Spatial Radius:11, Color Distance:8



Εικόνα 5.26 – Όρια πολυγώνων, κατάτμησης, με συντελεστές Spatial Radius:11, Color Distance:8



Εικόνα 5.27- Κατάτμηση εικόνας, με συντελεστές $Spatial\ Radius:8$, $Color\ Distance:5.5$



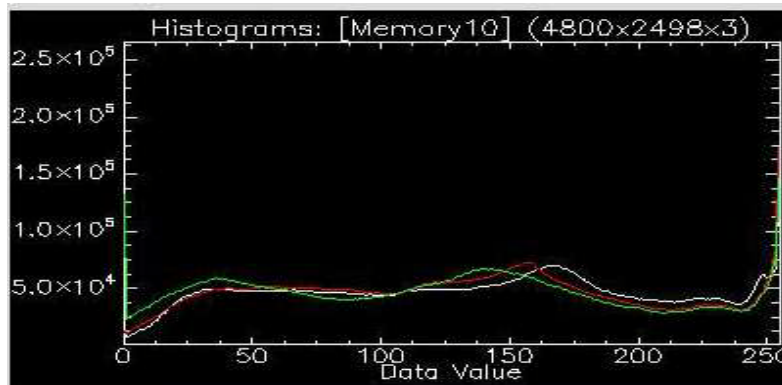
Εικόνα 5.28 Όρια πολυγώνων, κατάτμησης με συντελεστές $Spatial\ Radius:8$, $Color\ Distance:5.5$

5.4 Ο.Α.Κ.Α. Ολυμπιακές Εγκαταστάσεις, Αθήνα

Οι ολυμπιακές εγκαταστάσεις στην Αθήνα, αποτελούν κάποιες από τις πιο γρήγορα και δραστικά διαφοροποιημένες εκτάσεις στην Αθήνα. Η πιο κεντρική από αυτές το συγκρότημα ΟΑΚΑ, είναι το σημείο εφαρμογής, του αλγορίθμου στο τέταρτο πείραμα.



Εικόνα 5.29 – Αρχικά Δεδομένα

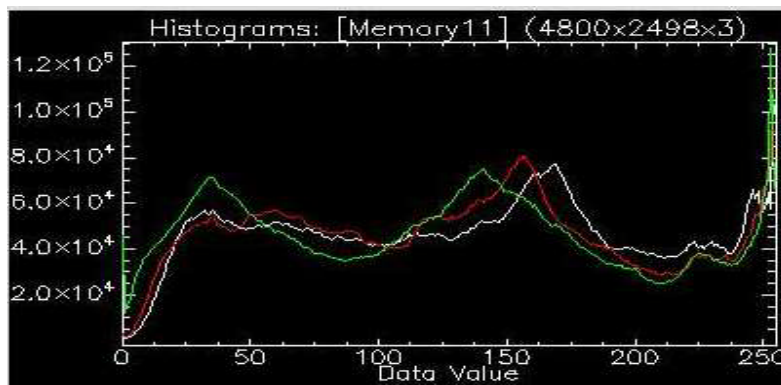


Εικόνα 5.30

Το ιστόγραμμα, υποδεικνύει την τάση ομογενοποίησης, που διακατέχει τα περισσότερα πυκνοκατοικημένα, αστικά τοπία.



Εικόνα 5.31- Κατατεμημένη εικόνα με συντελεστές Spatial Radius:11, Color Distance:8

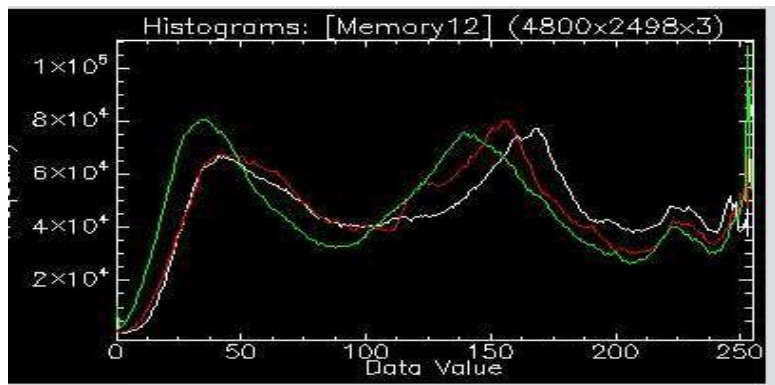


Εικόνα 5.32 – Ιστόγραμμα κατάτμησης με συντελεστές Spatial Radius:11, Color Distance:8

Μετά την εφαρμογή του αλγορίθμου, mean shift, φαίνεται σημαντική διαφορά στο ιστόγραμμα. Τα πιθανά κέντρα συσσωρεύσεων είναι πλέον πιο διακριτά, ωστόσο δεν έχουν ομαλοποιηθεί, ακόμα, οι πεδιάδες στο ιστόγραμμα, που σημαίνει ότι οι συντελεστές που έχουν χρησιμοποιηθεί θα πρέπει να αλλάξουν.

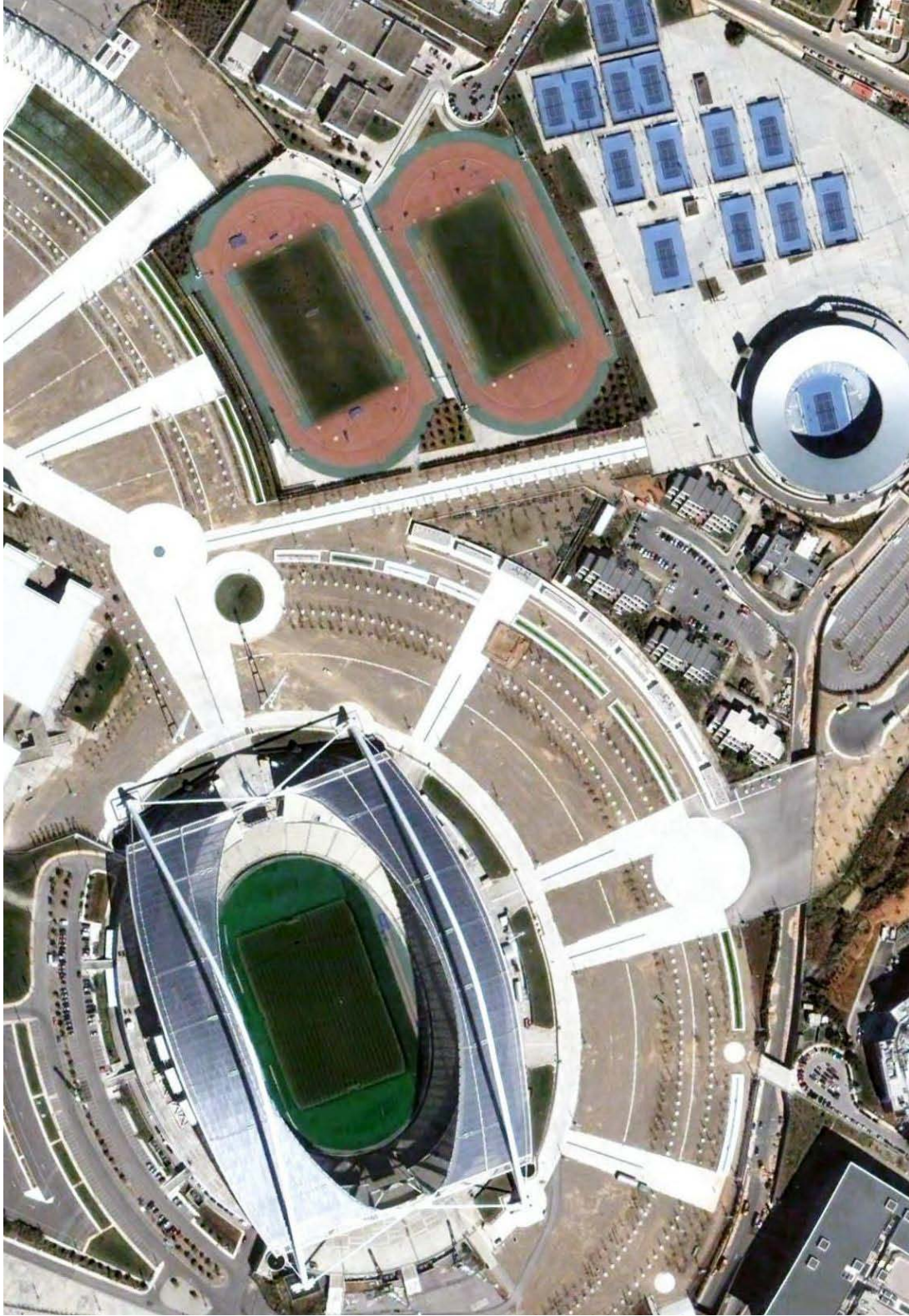


Εικόνα 5.33- Spatial Radius:13, Color Distance:8



Εικόνα 5.34

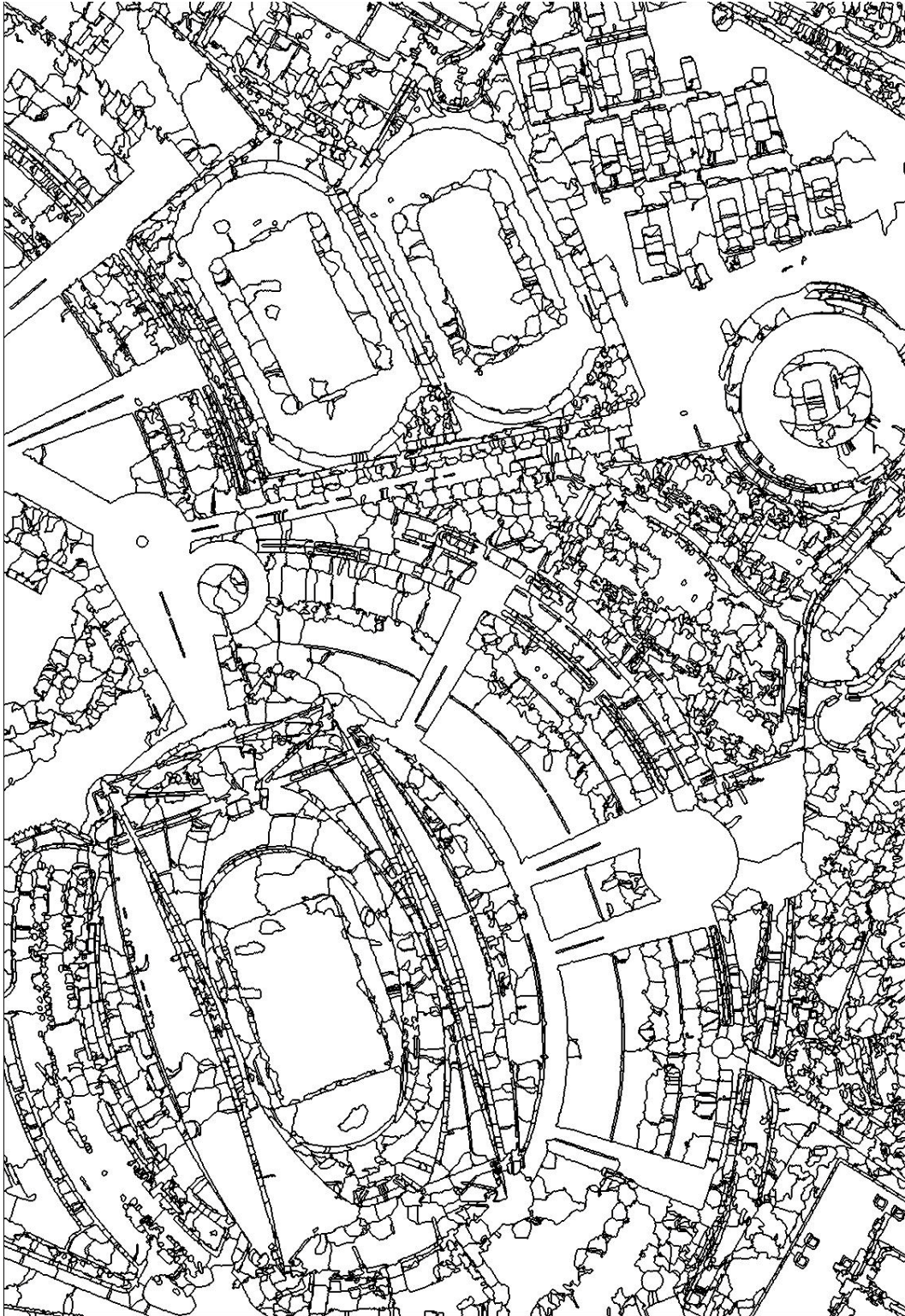
Με αυτό το συνδυασμό οι κορυφές είναι πλέον φανερά διαφοροποιημένες, και μπορούν να χρησιμοποιηθούν με αυξημένη βεβαιότητα σαν κέντρα συσσωρεύσεων. Σε μεγέθυνση φαίνεται σε μεγαλύτερη λεπτομέρεια η ανάλυση εικόνας που πέτυχε ο αλγόριθμος.



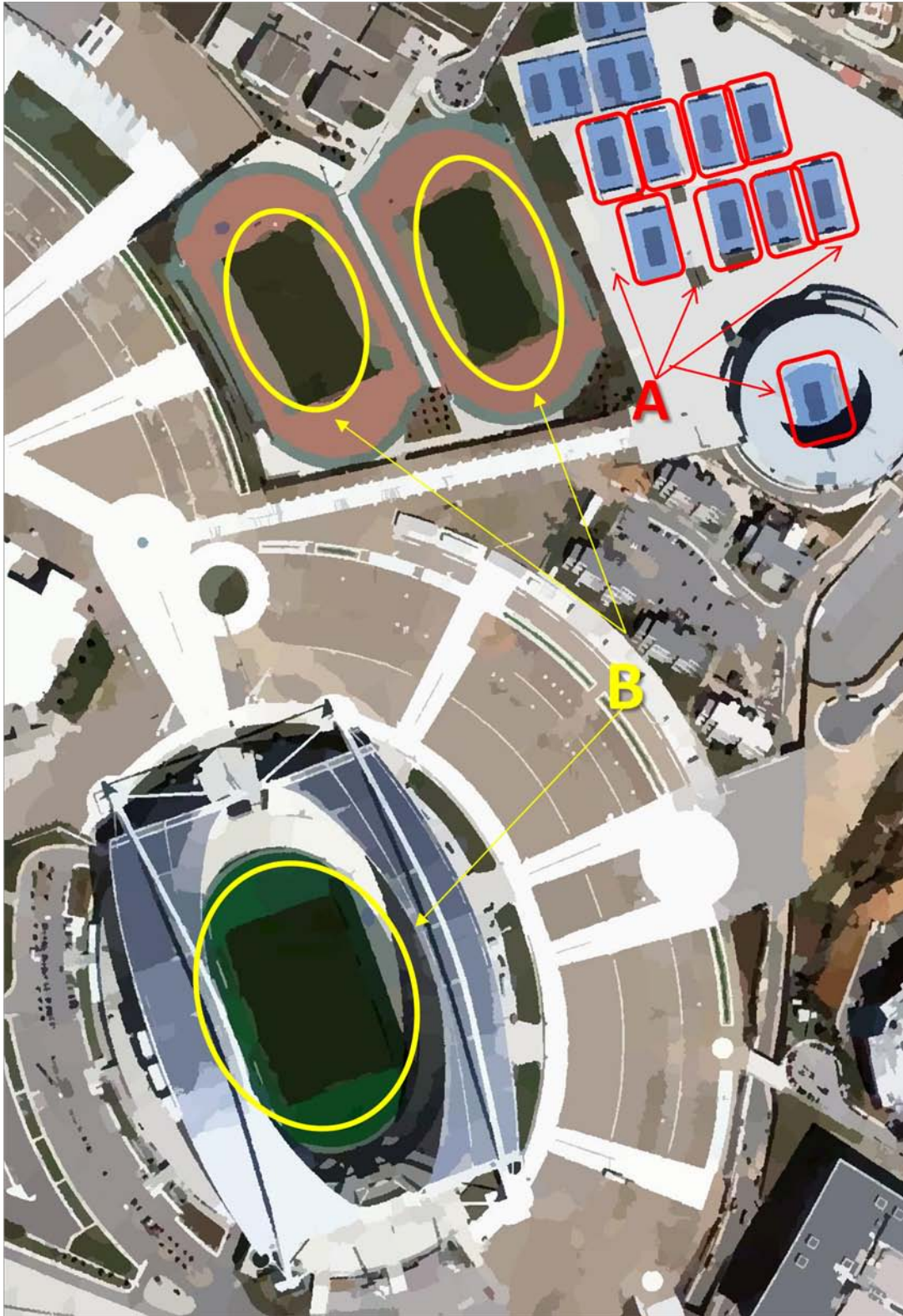
Εικόνα 5.35– Αρχικά δεδομένα



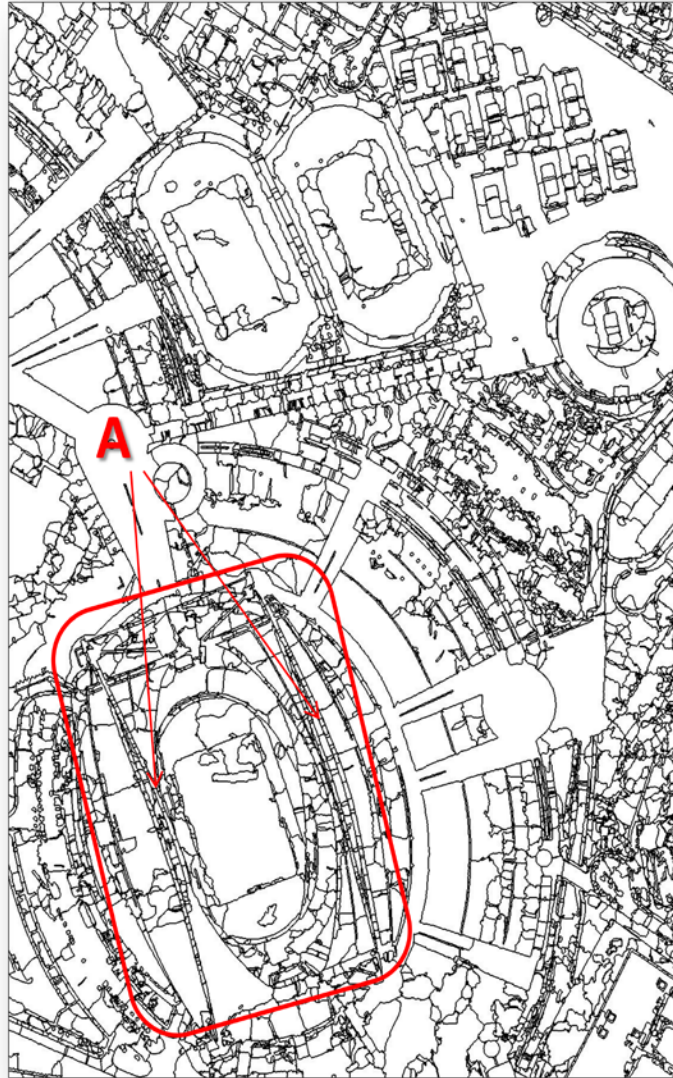
Εικόνα 5.36- Spatial Radius:11, Color Distance:8



Εικόνα 5.37- Όρια πολυγώνων, κατάτμησης Spatial Radius:11, Color Distance:8



Εικόνα 5.38 Κατάτμηση εικόνας με συντελεστές Spatial Radius:13, Color Distance:8



Εικόνα 5.39- Όρια πολυγώνων, κατάτμησης με συντελεστές Spatial Radius:13, Color Distance:8

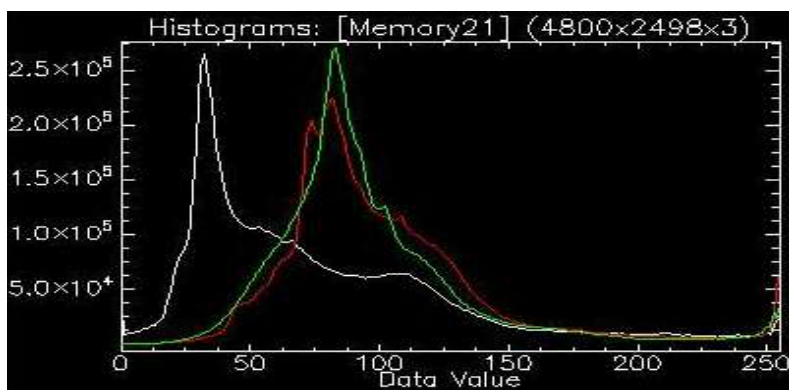
Όπως φαίνεται στην εικόνα [5.38] οι κεντρικές οντότητες της εικόνας, αρχίζουν και περιγράφονται σχετικά άρτια από την κατάτμηση. Τα κολυμβητήρια [5.38.A], το γήπεδο [5.38.B] κτλ, έχουν πλαισιωθεί είτε με ένα είτε με δύο τμήματα από την κατάτμηση, κάτι που θα βοηθούσε σημαντικά μια επόμενη κατηγοριοποίηση της εικόνας. Από την άλλη, πολυσύνθετες οντότητες, όπως το στάδιο, δεν έχει κατατμηθεί επαρκώς μιας και από τα όρια πολυγώνων [5.39A] τα τμήματα που έχουν δημιουργηθεί, δημιουργούν ένα κορεσμό, χωρίς να μπορούν να αντιστοιχίσουν συγκεκριμένες οντότητες.

5.5 Goa, Δυτική Ινδία

Η παραλία της Goa, στη Δυτική μεριά της ινδικής χερσονήσου, εμφανίζει μια έντονη ποικιλία εδαφοκάλυψης, μιας και περιέχει οικισμούς, καλλιεργήσιμες εκτάσεις, χέρσα χωράφια, μικρά δασύλλια, κα όλα αυτά ανάμεσα σε μια μεγάλου μήκους ακτογραμμή, και τον ομώνυμο ποταμό.



Εικόνα 5.40– Αρχικά δεδομένα

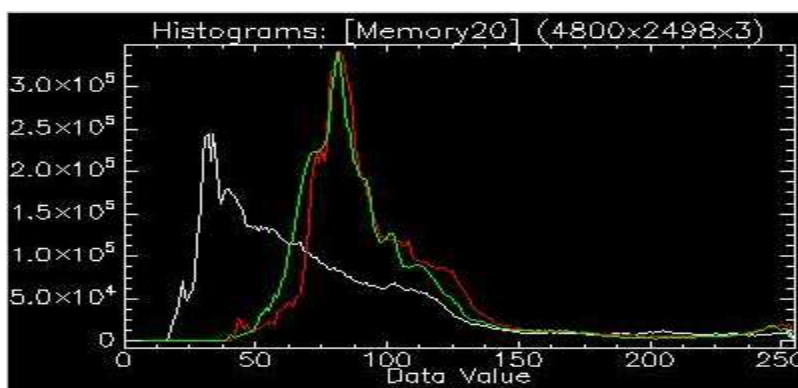


Εικόνα 5.41 - Ιστόγραμμα αρχικών δεδομένων

Το ιστόγραμμα υποδεικνύει ότι υπάρχει χώρος για clustering, με σχετικά μεγάλο ποσοστό βεβαιότητας, αν και η θάλασσα σε συνδυασμό με το ποτάμιο σύστημα, επηρεάζει πολύ την κατανομή των χρωματικών τιμών των εικονοστοιχείων.



Εικόνα 5.42 Κατάτμηση εικόνας με συντελεστές: Spatial Radius:10, Color Distance:20



Εικόνα 5.43 – Ιστόγραμμα κατατετημένης εικόνας με συντελεστές: Spatial Radius:10, Color Distance:20

Το ιστόγραμμα [5.43] αρχίζει και τροποποιείται ακολουθώντας τα clusters που δημιουργηθήκαν για την κατάτμηση της εικόνας. Όπως φαίνεται και στην μεγέθυνση των αποτελεσμάτων [5.44, 5.45, 5.46], το πράσινο και μπλε χρώμα δημιουργεί το φαινόμενο κηλίδας (blurring effect) κάτι που εξηγείται και από την έλλειψη σημαντικών δευτερευόντων κορυφών στο προηγθέν ιστόγραμμα [5.41].

Ακολουθούν τρεις διαφορετικές προσεγγίσεις για τη μείωση του συγκεκριμένου φαινομένου.



Εικόνα 5.44 Κατάτμηση εικόνας με συντελεστές: Spatial Radius:10, Color Distance:25



Εικόνα 5.45 Κατάτμηση εικόνας με συντελεστές: Spatial Radius:15, Color Distance:15



Εικόνα 5.46 Κατάτμηση εικόνας με συντελεστές: Spatial Radius:13, Color Distance:8

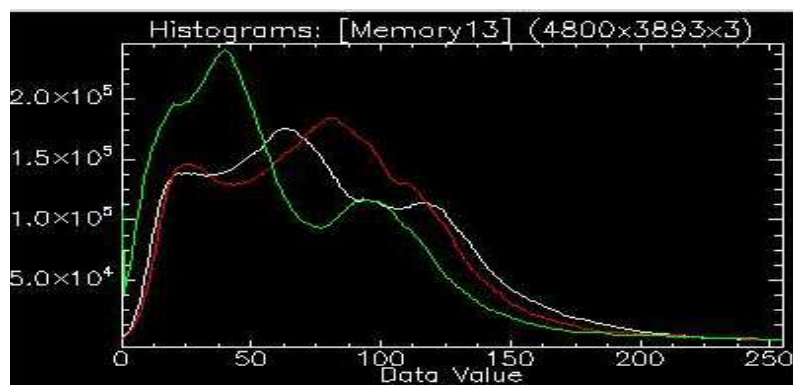
Όπως γίνεται κατανοητό από την τελευταία εικόνα, η σημαντική μείωση της παραμέτρου Color Distance, δείχνει να διορθώνει λίγο το φαινόμενο της κηλίδας, χωρίς όμως να το εξαλείφει. Η συνεχόμενη μείωση όμως, ήδη έχει αρχίσει να επηρεάζει τον εντοπισμό των οικισμών στη συγκεκριμένη περιοχή. Σε αυτή την εφαρμογή, κάποιοι χωρικοί περιορισμοί, η δυνατότητα αυτόματης αναπροσαρμογής του αλγορίθμου mean shift ή μια ανάλυση σε πολλά επίπεδα, θα αναβάθμιζαν σημαντικά την ποιότητα του αποτελέσματος.

5.6 Εδιμβούργο, Ηνωμένο Βασίλειο

Η περίπτωση του Εδιμβούργου είναι μια κλασική εφαρμογή του αλγορίθμου κατάτμησης, δεδομένης της ομοιότητας των σπιτιών, στις αγγλικές γειτονιές.

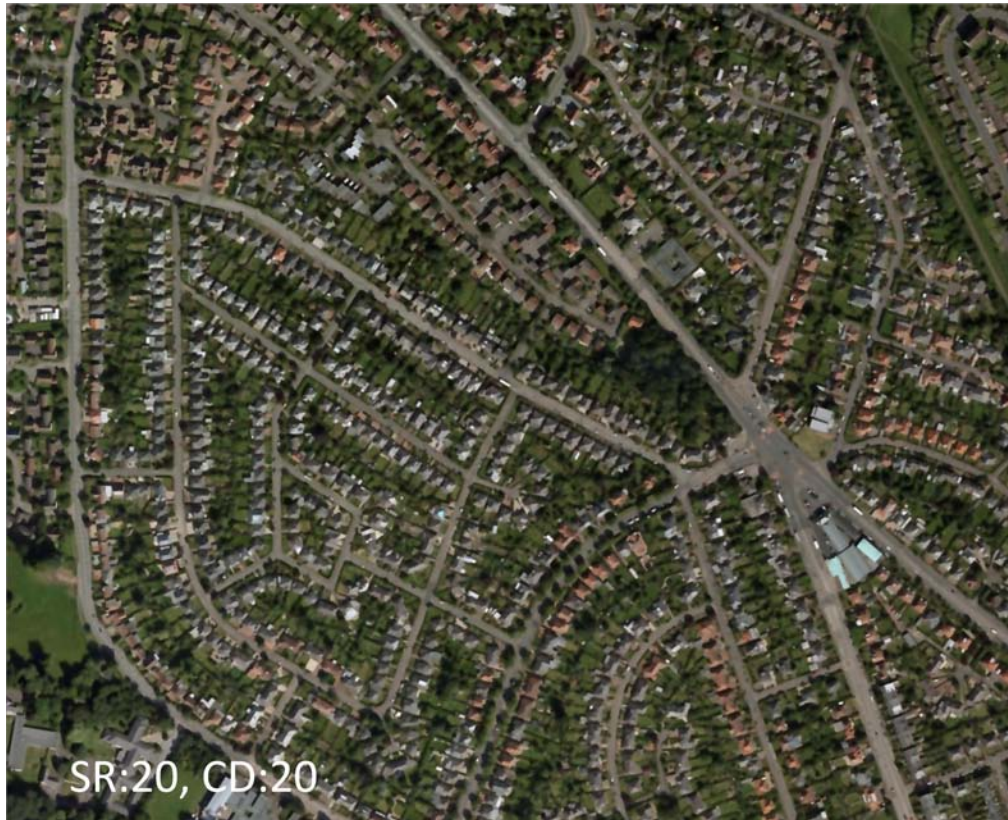


Εικόνα 5.47– Αρχικά δεδομένα

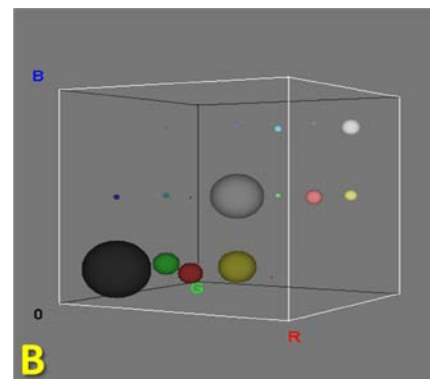
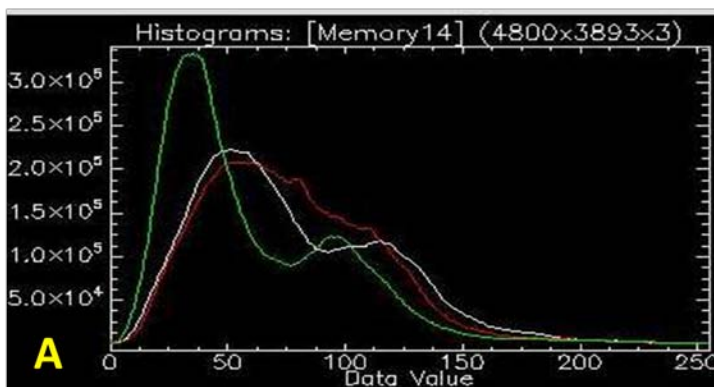


Εικόνα 5.48 – Ιστόγραμμα Αρχικών Δεδομένων

Η προδιάθεση του ιστογράμματος είναι εμφανής. Οι κορυφές δείχνουν ότι υπάρχουν εν δυνάμει σημεία για κέντρα συσσωρεύσεων. Αυτό θα διευκολύνει πολύ τον αλγόριθμο, να δημιουργήσει τμήματα πάνω στην εικόνα, τα οποία θα μπορούν να περιγραφούν λεκτικά.



Εικόνα 5.49 Κατάτμηση εικόνας με συντελεστές: Spatial Radius:20, Color Distance:20



Εικόνα 5.50 Ιστογράμμο κατατετημένης εικόνας με συντελεστές: Spatial Radius:20, Color Distance:20 και 3D απόδοση των κέντρων συσσωρεύσεων που επιλέχθηκαν κατά την κατάτμηση

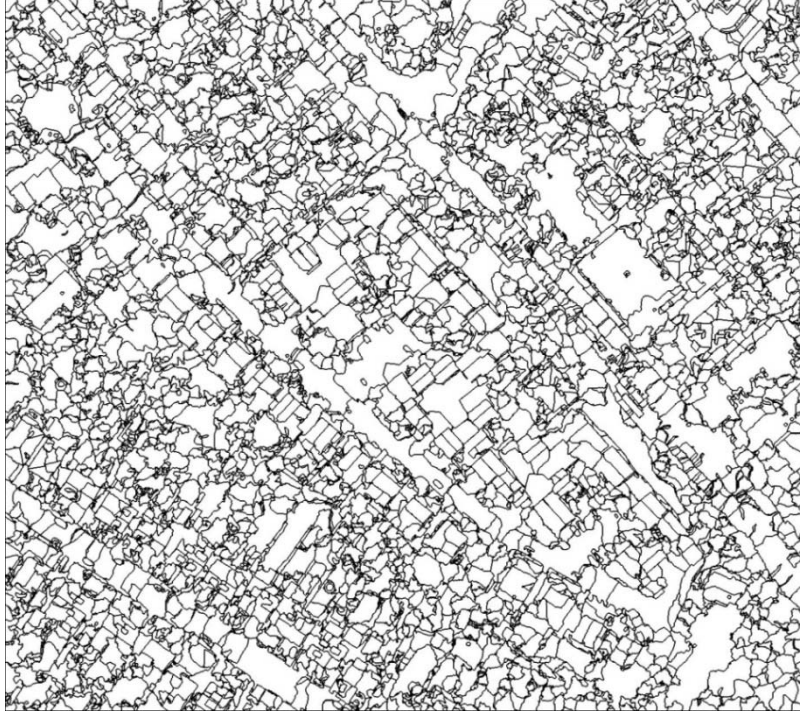
Μετά τις πρώτες εφαρμογές του αλγορίθμου, είναι απολύτως προφανές, ότι η διαδικασία mean shift ακολουθεί την ευρύτερη φιλοσοφία του ιστογράμματος. Στην 3D εικόνα των κέντρων συσσωρεύσεων [5.50B], έχουμε μια καλύτερη απεικόνιση του ίδιου ιστογράμματος, στον RGB χρωματικό χώρο. Ακολουθούν σε μεγέθυνση τα αποτελέσματα της κατάτμησης.



Εικόνα 5.51– Αρχικά δεδομένα



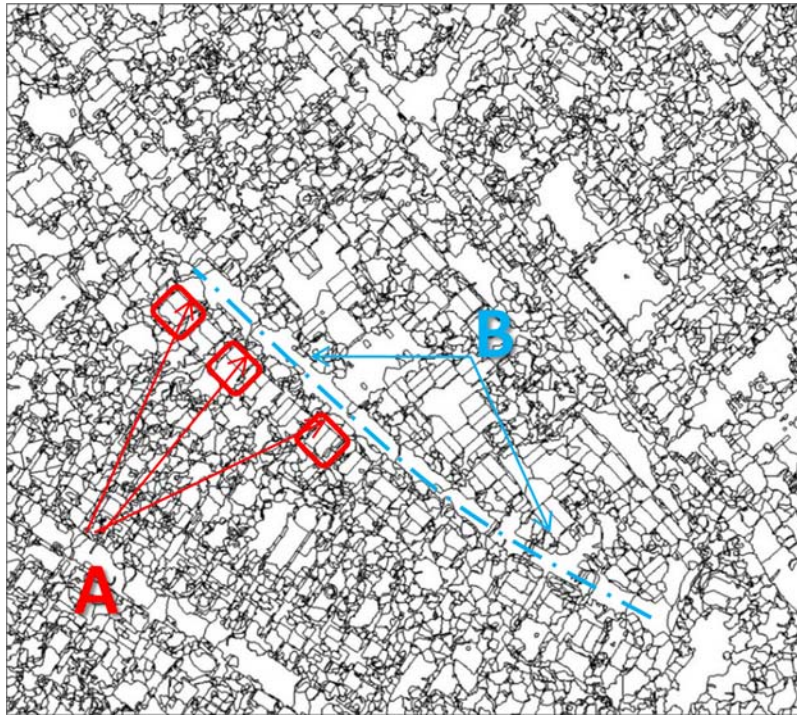
Εικόνα 5.52 - Κατάτμηση εικόνας με συντελεστές: Spatial Radius:11, Color Distance:8



Εικόνα 5.53 Όρια πολυγώνων, προηγούμενης κατάτμησης με συντελεστές: Spatial Radius:11, Color Distance:8



Εικόνα 5.54 Κατάτμηση εικόνας με συντελεστές: Spatial Radius:13, Color Distance:9



Εικόνα 5.55 Όρια πολυγώνων, κατάτμησης με συντελεστές: Spatial Radius:13, Color Distance:9

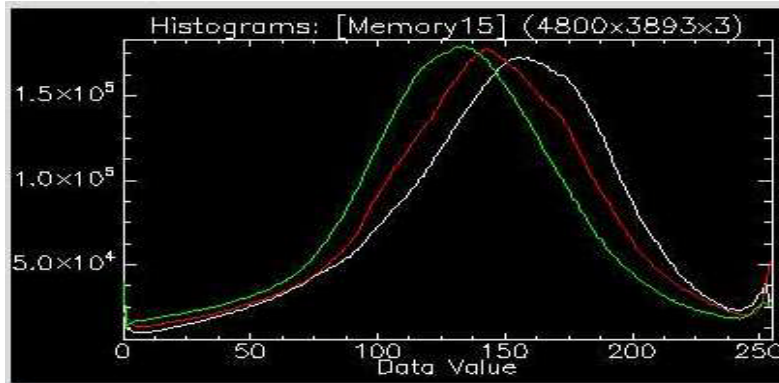
Η εφαρμογή του αλγορίθμου εδώ, είναι ικανοποιητική. Καθώς οι στέγες των σπιτιών [5.55A] διαφοροποιούνται σημαντικά τόσο από το πράσινο της περιοχής, όσο και από το οδόστρωμα [5.55B].

5.7 West Ghetto, Ναϊρόμπι

Στην προσπάθεια μείωσης εγκληματικότητας, σε γκέτο όπως της λατινικής Αμερικής, ή της Αφρικής, είναι απαραίτητο να γίνει καταγραφή των υπάρχοντων καταυλισμών, να υπάρξει κάποια δυνατότητα εντοπισμού σε διαφορές κατά την πάροδο του χρόνου κτλ. Στη συγκεκριμένη εφαρμογή, θα μεταφερθούμε σε ένα από τα πληθυσμιακά μεγαλύτερα γκέτο της Αφρικής. Η αναρχία επικρατεί σε όλη την εικόνα, κάνοντας το έργο του αλγόριθμου, σχεδόν αδύνατο. Για τον καλύτερο εντοπισμό οικιών, θα πρέπει αφενός να χρησιμοποιηθεί μικρή χρωματική απόσταση, για να μην υπάρχει χρωματική σύγχυση σε γειτονικές οικίες, και αφετέρου θα πρέπει να υπάρξει έντονος πειραματισμός με τη χωρική ακτίνα που θα χρησιμοποιηθεί. Ακολουθούν κάποια από τα αποτελέσματα της εφαρμογής.



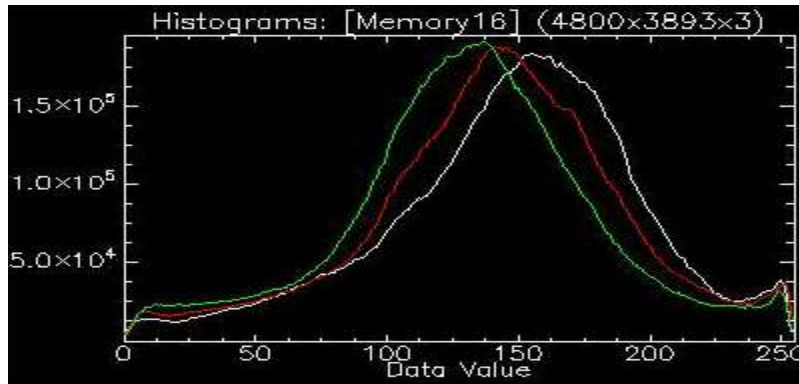
Εικόνα 5.56– Αρχικά δεδομένα



Εικόνα 5.57 Ιστόγραμμα αρχικών δεδομένων



Εικόνα 5.58 Κατατεμημένη εικόνα με συντελεστές: Spatial Radius:15, Color Distance:15

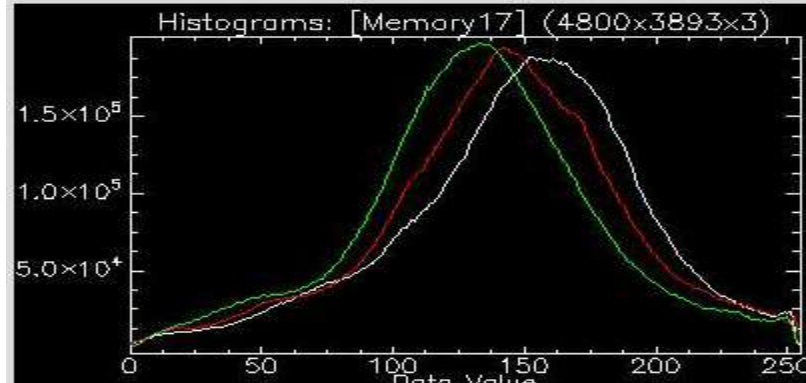


Εικόνα 5.59 Ιστόγραμμα κατατετημμένης εικόνας με συντελεστές: Spatial Radius:15, Color Distance:15

Παρατηρείται ότι ο αλγόριθμος, προσπαθεί να δημιουργήσει κάποιου είδους ανωμαλία για τοποθέτηση των κέντρων συσσωρεύσεων, ωστόσο ο έντονος γκαουσιανός χαρακτήρας των καμπυλών, δεν επιτρέπει πολλές βελτιώσεις.



Εικόνα 5.60 Κατατετημμένη εικόνα με συντελεστές: Spatial Radius:20, Color Distance:20



Εικόνα 5.61 Ιστόγραμμα κατατετημένης εικόνας με συντελεστές: Spatial Radius:20, Color Distance:20

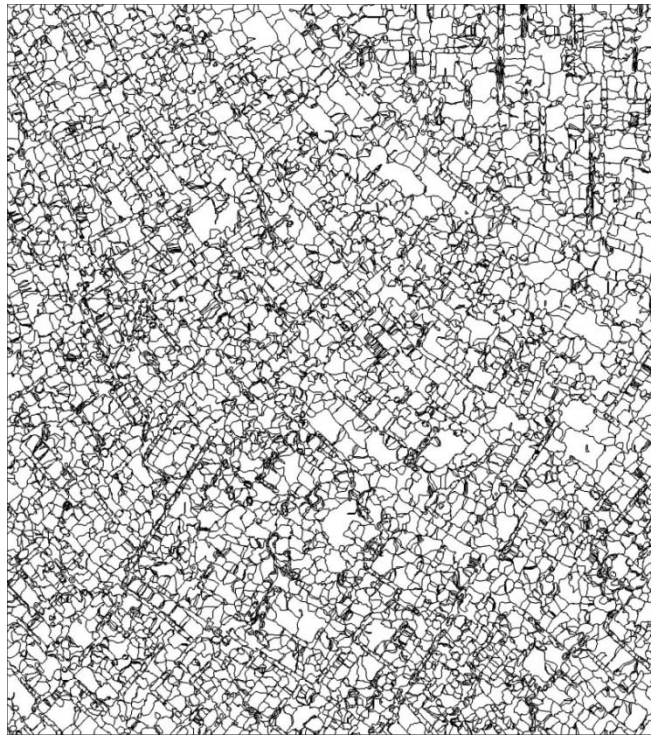
Ακριβώς όπως δείχνει το ιστόγραμμα [5.61], φαίνεται ότι η χρωματική σύγχυση δεν μπορεί να αποφευχθεί δεδομένης της ομοιότητας του εδάφους με αρκετές από της στέγες των οικιών



Εικόνα 5.62 Αρχικά Δεδομένα



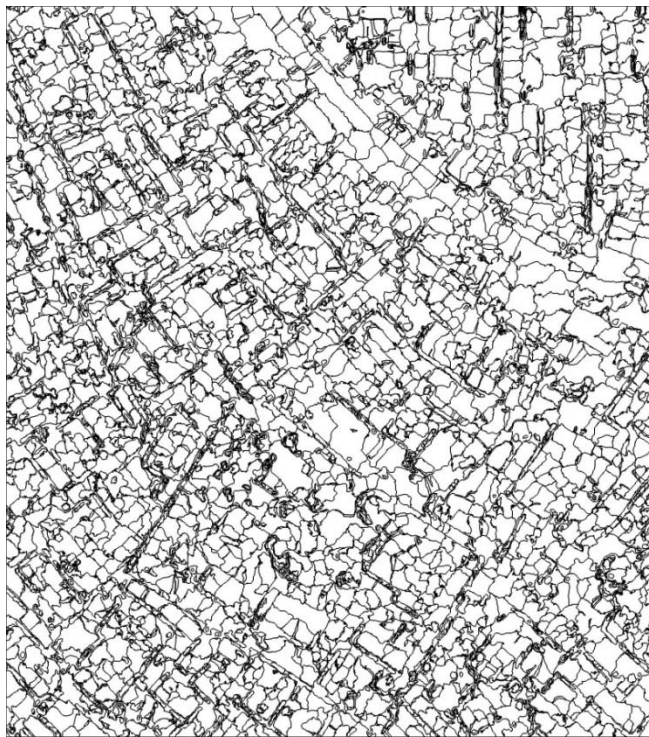
Εικόνα 5.63- Κατατετημένης εικόνας με συντελεστές: Spatial Radius:11, Color Distance:8



Εικόνα 5.64 – Ιστόγραμμα κατατετημένης εικόνας με συντελεστές: Spatial Radius:11, Color Distance:8



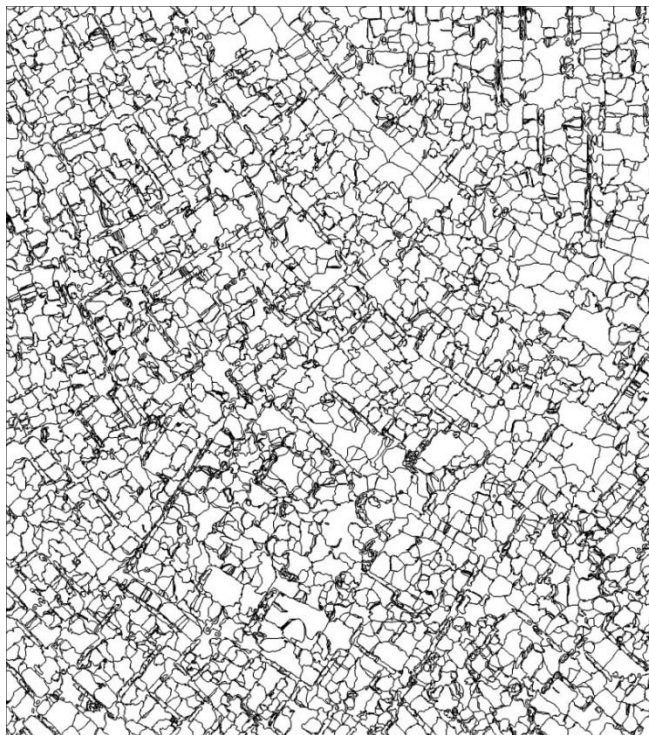
Εικόνα 5.65- Κατατεμημένη εικόνα με συντελεστές: Spatial Radius:17, Color Distance:7



Εικόνα 5.66 Όρια πολυγώνων, κατατεμημένης εικόνας με συντελεστές: Spatial Radius:17, Color Distance:7

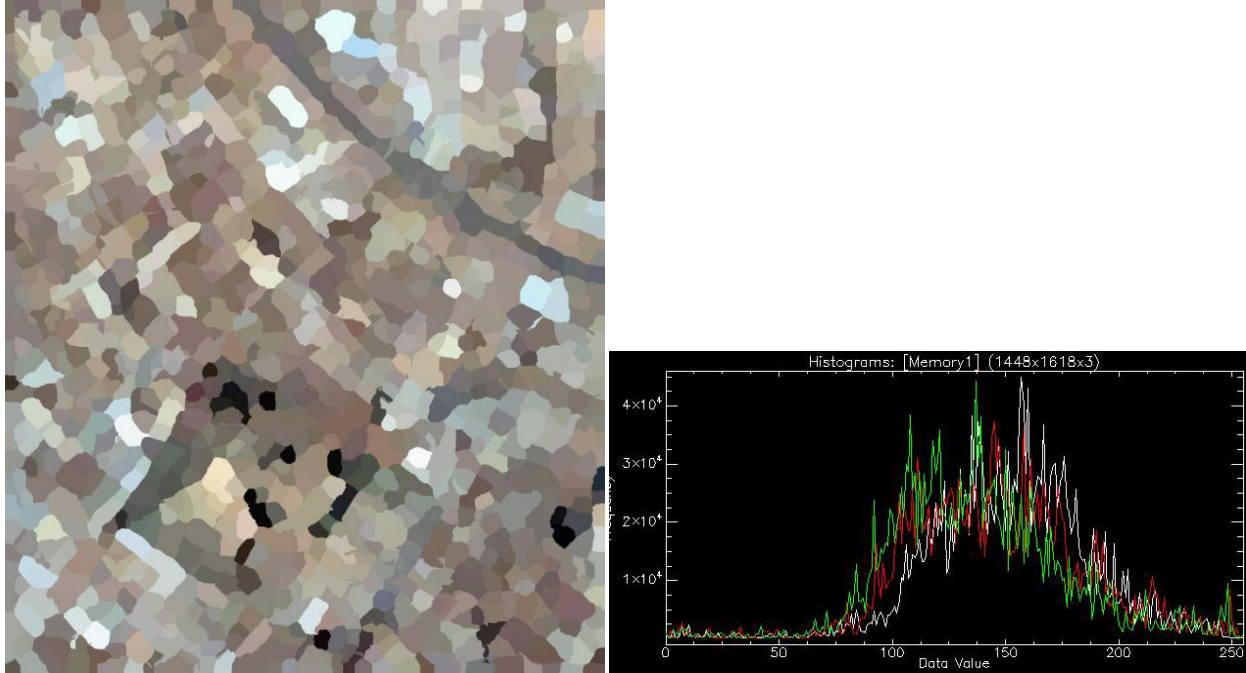


Εικόνα 5.67- Κατατεμημένη εικόνα με συντελεστές: Spatial Radius:15, Color Distance:9



Εικόνα 5.68 Όρια πολυγώνων, Κατατεμημένης εικόνας με συντελεστές: Spatial Radius:15, Color Distance:9

Στις εικόνες [5.66] & [5.67] φαίνεται ο κορεσμός του προϊόντος της κατάτμησης. Σε μια προσπάθεια, να αποκτήσει μια μεγαλύτερη σαφήνεια η ανάλυση της εικόνας, έγινε μια εφαρμογή με ιδιαίτερα αυξημένους τους συντελεστές της mean shift ανάλυσης. Το αποτέλεσμα, αν και αφαιρετικό, δίνει καλύτερη αίσθηση της υπάρχουσας κατάστασης.



5.68α. - Κατατετημένη εικόνα και ιστόγραμμα με συντελεστές: Spatial Radius:20, Color Distance:60

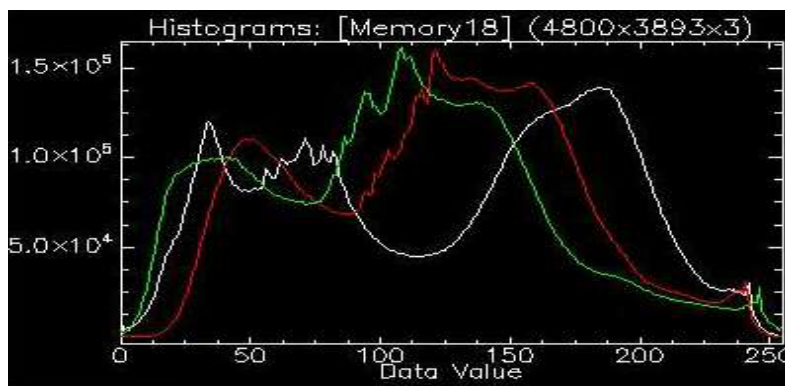
Αν και η τελευταία προσέγγιση είναι εμφανώς πιο ακραία – φαίνεται καθαρά και από το ιστόγραμμα [5.68B], ότι τα κέντρα συσσωρεύσεων μπορούν πλέον να υποβοηθήσουν την υπολογιστική όραση στον εντοπισμό οντοτήτων, οι οποίες μπορούν να ενταχθούν σε συγκεκριμένο σημασιολογικό πλαίσιο.

5.8 Πόλη της Βενετίας, Ιταλία

Η Βενετία, αποτελεί σε ένα μεγάλο βαθμό το ιδανικό πεδίο εφαρμογής της κατάτμησης, μιας και τόσο οι χρωματικές διαφορές, όσο και τα σχήματα των αντικειμένων, καταστούν το εγχείρημα τυπική υπόθεση. όπως αναδεικνύουν και τα αποτελέσματα, ο αλγόριθμος σε αυτή την εφαρμογή, έχει αυξημένη αποδοτικότητα.



Εικόνα 5.69– Αρχικά δεδομένα

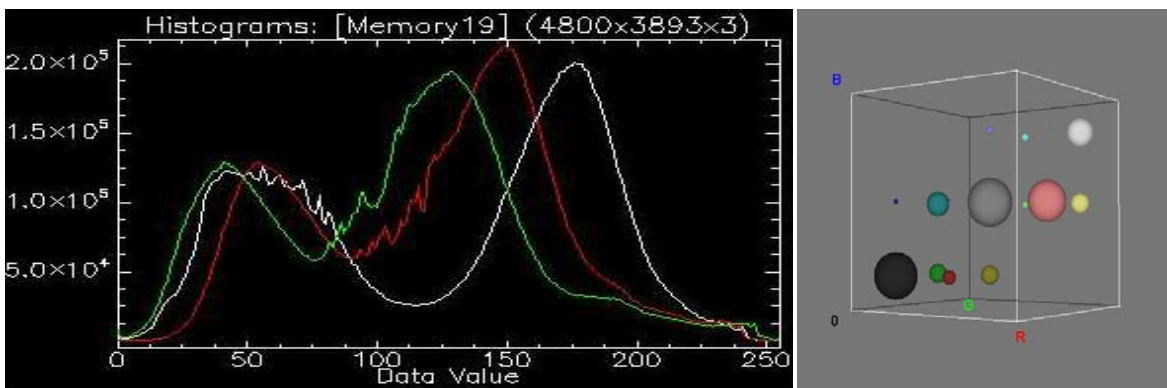


Εικόνα 5.70 – Ιστόγραμμα αρχικών δεδομένων

Το ιστόγραμμα των αρχικών δεδομένων, δείχνει ότι τα clusters, είναι ήδη σχεδόν διαμορφωμένα, συνεπώς ο αλγόριθμος οφείλει να ακολουθήσει τις υπάρχουσες καμπύλες.



Εικόνα 5.71 Κατατεμημένη εικόνα με συντελεστές: Spatial Radius:20, Color Distance:20



Εικόνα 5.72 - Ιστόγραμμα κατατεμημένης εικόνας με συντελεστές: Spatial Radius:20, Color Distance:20 και 3D απόδοση των κέντρων συσσωρεύσεων που επιλέχθηκαν κατά την κατάτμηση

Ακριβώς όπως αναμενόταν, η δημιουργία των ανάλογων clusters, μείωσε τις διακυμάνσεις στις χαμηλές ψηφιακές τιμές των εικονοστοιχείων, κρατώντας όμως τις απαραίτητες κορυφές σαν κέντρα συσσωρεύσεων. Στο δεξί κομμάτι του ιστογράμματος, φαίνεται πιο καθαρά η διασπορά των clusters που δημιουργηθήκαν στο χώρο RGB.



Εικόνα 5.73- Κατατεμημένη εικόνα με συντελεστές: Spatial Radius:7, Color Distance:6.5



Εικόνα 5.74 – Όρια πολυγώνων κατατεμημένης εικόνας με συντελεστές: Spatial Radius:7, Color Distance:6.5



Εικόνα 5.75 – Μια πιο λεπτομερής επίδειξη της κατάτμησης με συντελεστές: Spatial Radius:7, Color Distance:6.5



Εικόνα 5.76 –Όρια πολυγώνων, στην κατάτμηση με συντελεστές: Spatial Radius:7, Color Distance:6.5

5.9 Ηράκλειο, Κρήτη

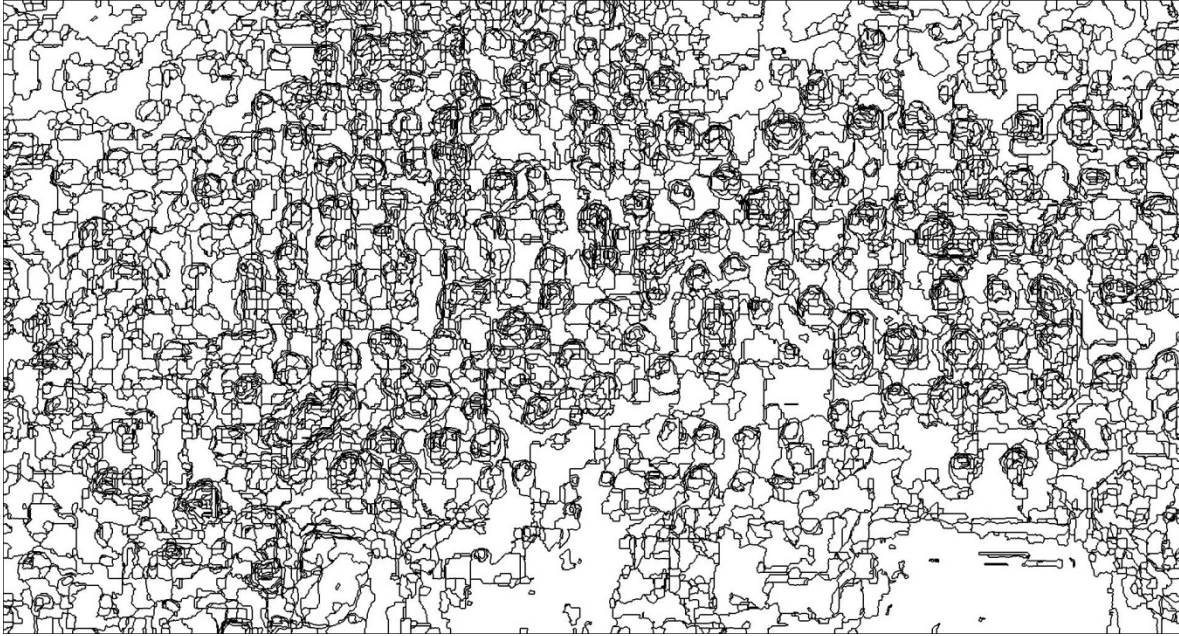
Στο παρελθόν, μια από τις πιο χρήσιμες, εφαρμογές της κατάτμησης εικόνας, στην Ελλάδα, ήταν ο εντόπισμός και η καταμέτρηση ριζών ελιάς. Σε αυτή την εικόνα, έχουμε ένα προάστειο του ηρακλείου, Κρήτης, στο οποίο δοκιμάζεται ο αλγόριθμος του λογισμικού.



Εικόνα 5.77 – Αρχικά Δεδομένα



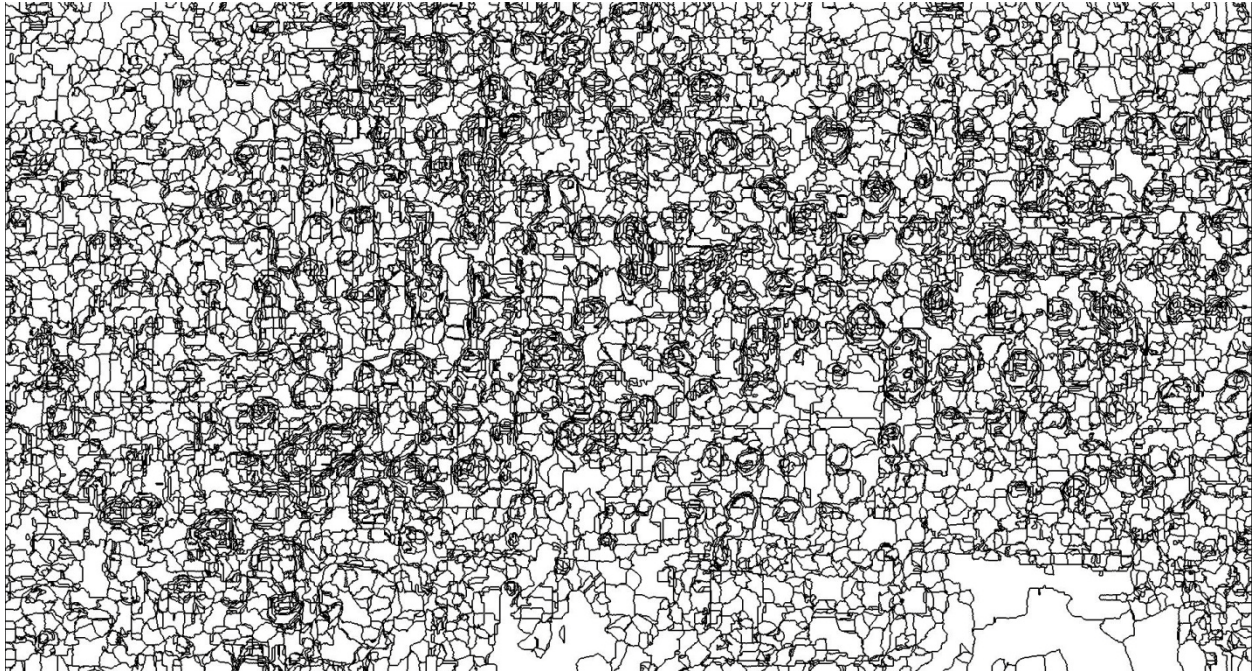
Εικόνα 5.78- Κατάτμηση εικόνας με συντελεστές: Spatial Radius:9, Color Distance:2



Εικόνα 5.79 Όρια πολυγώνων, κατάτμησης με συντελεστές:- Spatial Radius:9, Color Distance:2



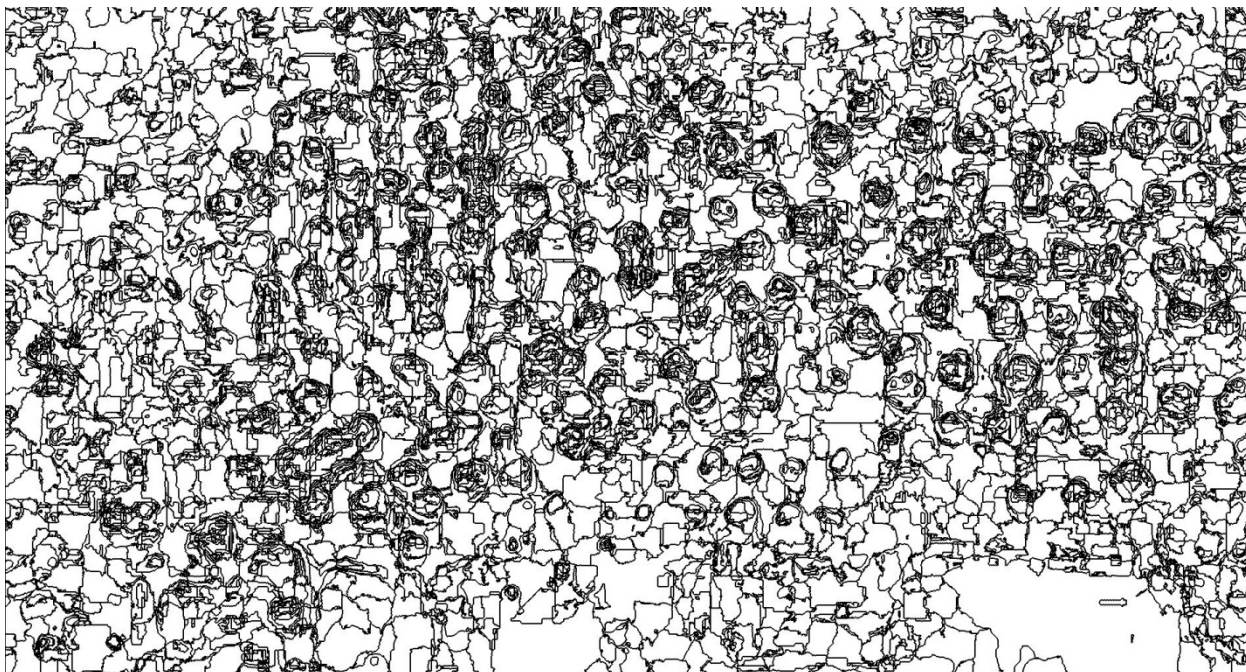
Εικόνα 5.80- Κατάτμηση εικόνας με συντελεστές:2Spatial Radius:10, Color Distance:2



Εικόνα 5.81- Όρια πολυγώνων, κατάτμησης εικόνας με συντελεστές: Spatial Radius:10, Color Distance:2



Εικόνα 5.82- Κατάτμηση εικόνας με συντελεστές: Spatial Radius:20, Color Distance:2



Εικόνα 5.83- Όρια πολυγώνων, κατάτμησης με συντελεστές *Spatial Radius:20, Color Distance:2*

Κρατώντας σταθερή σε χαμηλά επίπεδα την χρωματική απόσταση της συσσώρευσης [5.79, 5.81], η χωρική ακτίνα, είναι σε θέση να εντοπίσει τις περισσότερες ρίζες της εικόνας, κάνοντας χρήση ενός εύρους διαφορετικών τιμών. Τα πολύγωνα-προϊόντα της κατάτμησης είναι αρκετά συγκεχυμένα [5.82], αλλά μπορούν να αποτελέσουν το πρώτο βήμα μιας λεπτομερούς κατηγοριοποίησης της εικόνας. Είναι προφανές ότι για την καλύτερη αξιολόγηση μιας τέτοιας προσπάθειας, θα πρέπει να υπάρξει και έρευνα πεδίου, ωστόσο σαν πρώτη προσέγγιση φαίνεται να κατέχει μια αρκετά μεγάλη σταθερότητα στην απόδοσή του.

6 Αξιολόγηση - Συμπεράσματα – Μελλοντικές Εργασίες

6.1 Αξιολόγηση

- Για την αξιολόγηση του αλγορίθμου κατάτμησης που αναπτύχθηκε στην παρούσα πτυχιακή, χρησιμοποιήθηκε ένα εξειδικευμένο πακέτο ανάλυσης εικόνας. Το e-Cognition, είναι ένα software το οποίο μέσα στις άλλες δυνατότητες του, χρησιμοποιεί ανεπτυγμένους αλγορίθμους για πολυεπίπεδη κατάτμηση εικόνων.
- Για να επιτευχθεί μια σύγκριση μεταξύ των αποτελεσμάτων χρησιμοποιήθηκαν οι εξής κοινές αρχές:
 1. Η χωρική διάσταση της ανάλυσης που γίνεται και με τους δύο αλγορίθμους, θα είναι ίδια.
 2. Δεδομένου ότι το Earthwatch δίνει τη δυνατότητα μονοεπίπεδης κατάτμησης εικόνας, η χρήση του eCognition θα πρέπει και αυτή να περιοριστεί σε ένα επίπεδο τη φορά.
 3. Αντίστοιχα, λαμβάνοντας υπόψη ότι το eCognition δίνει τη δυνατότητα χρήσης κριτηρίων που δεν υπάρχουν στο Earthwatch, θα παρουσιάζονται στη σύγκριση δύο eCognition αποτελέσματα – ένα με αυστηρά χρωματική κατάτμηση, χωρίς κανένα άλλο κριτήριο και ένα δεύτερο αποτέλεσμα με τη χρήση συγκεκριμένης αναλογίας των εναλλακτικών κριτηρίων που δίνονται.
- Θα ήταν σημαντικό να σημειωθεί, ότι η σύγκριση των αποτελεσμάτων δίνεται με καθαρά ενδεικτικό χαρακτήρα, μιας και οι συντελεστές κριτηρίων που χρησιμοποιήθηκαν δεν είναι απαραίτητα οι βέλτιστοι για το περιβάλλον eCognition.

6.1.1 Αεροδρόμιο Ελ. Βενιζέλος, Αθήνα



EarthWatch_Segmentation

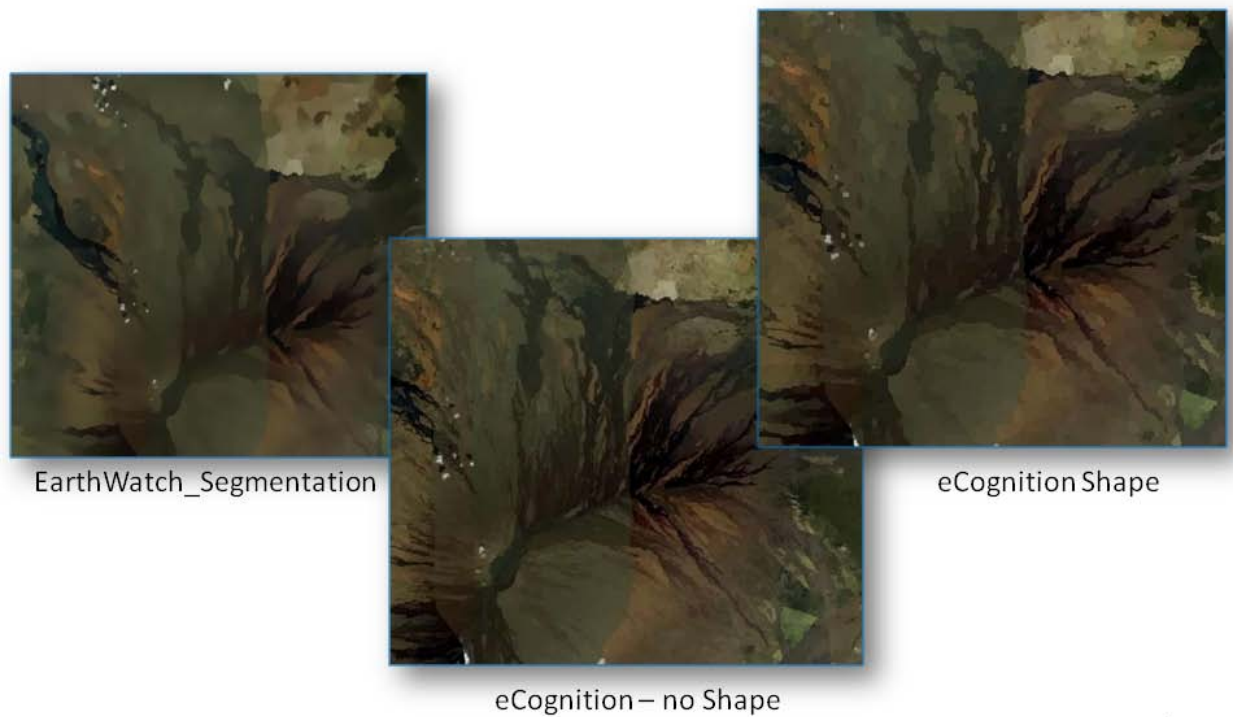
e-Cognition – no Shape

e-Cognition Shape

Εικόνα 6.1 Σύγκριση του Αλγορίθμου EarthWatch-MeanShift με δύο εναλλακτικές παραμέτρους στο eCognition

- Στη σύγκριση των δύο πρώτων εικόνων είναι προφανές ότι ο αλγόριθμος του eCognition θέλει περεταίρω τροποποίηση. Το γεγονός ότι στα κριτήρια της κατάτμησης δεν χρησιμοποιήθηκε καθόλου το σχηματικό κριτήριο, δίνει ένα αρκετά μειωμένης ποιότητας αποτέλεσμα. Με την εισαγωγή των σχηματικών κριτηρίων ωστόσο (compactness & smoothness) ακόμα και με προκαθορισμένες τιμές, παρατηρείται μια βελτίωση στις μεγάλες οντότητες του δείγματος (πύργος ελέγχου), και μια αισθητή υπεροχή στις μικρότερες (αυτοκίνητα) μιας και ο αλγόριθμος που περιλαμβάνεται στο πακέτο Earthwatch τα εμφανίζει συγκεχυμένα.

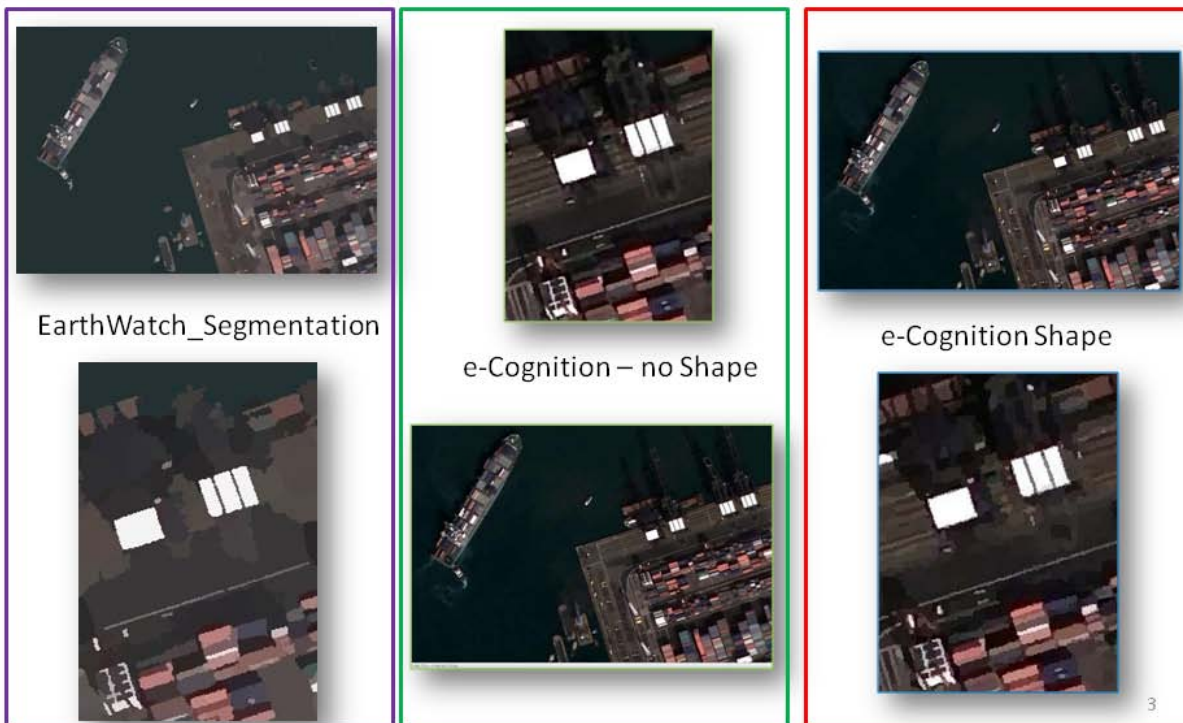
6.1.2 Big Island, Hawaii



Εικόνα 6.2 Σύγκριση του Αλγορίθμου EarthWatch-MeanShift με δύο εναλλακτικές παραμέτρους στο eCognition

- Σε αυτή την περίπτωση η ανάλυση στοχεύει σε εξαγωγή γεωμορφών, συνεπώς οι έννοιες Smoothness & Compactness δεν θα έπρεπε να αλλάζουν αισθητά το αποτέλεσμα το eCognition. Ωστόσο παρατηρείται ότι χωρίς αυτά η κατάτμηση υπολείπεται τόσο σε ποιότητα, ώστε οριακά να μπορεί να χρησιμοποιηθεί για περαιτέρω κατηγοριοποίηση. Στις 2 άλλες εικόνες τώρα, τα σύννεφα εξάχθηκαν σωστά, ενώ ο αλγόριθμος του Earthwatch δείχνει να πετυχαίνει κατά μέσο όρο, μεγαλύτερα τμήματα σαν αποτέλεσμα κατάτμησης, χωρίς ταυτόχρονα να χάνει χρήσιμη πληροφορία.

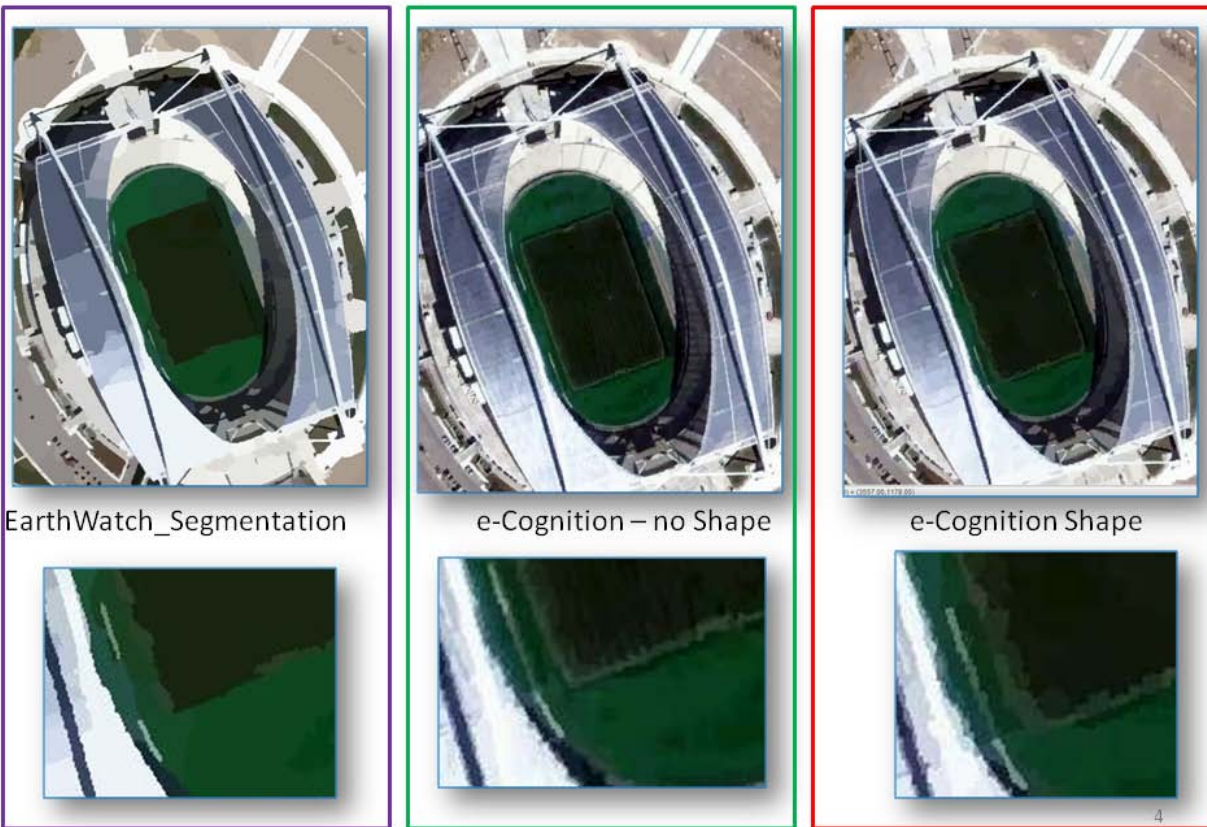
6.1.3 Victoria Harbour, Hong Kong



Εικόνα 6.3 Σύγκριση του Αλγορίθμου EarthWatch-MeanShift με δύο εναλλακτικές παραμέτρους στο eCognition

- Η δυνατότητα στάθμισης των αποτελεσμάτων με τη χρήση σχηματικών κριτηρίων δίνει ένα μεγάλο πλεονέκτημα στο eCognition. Τα containers, που έχουν το ίδιες χρωματικές αποχρώσεις μπαίνουν στην ίδια κατηγορία, οι σκιές στην εικόνα, περιορίζουν σημαντικά την αποτελεσματικότητα της κατάτμησης, η θάλασσα εξάγεται ικανοποιητικά και στις τρεις εικόνες, ωστόσο το Earthwatch χάνει κάποια τμήματα πληροφορίας, κυρίως λόγω έλλειψης της ικανότητας να συμπεριλάβει σχηματική παράμετρο στην κατάτμηση.

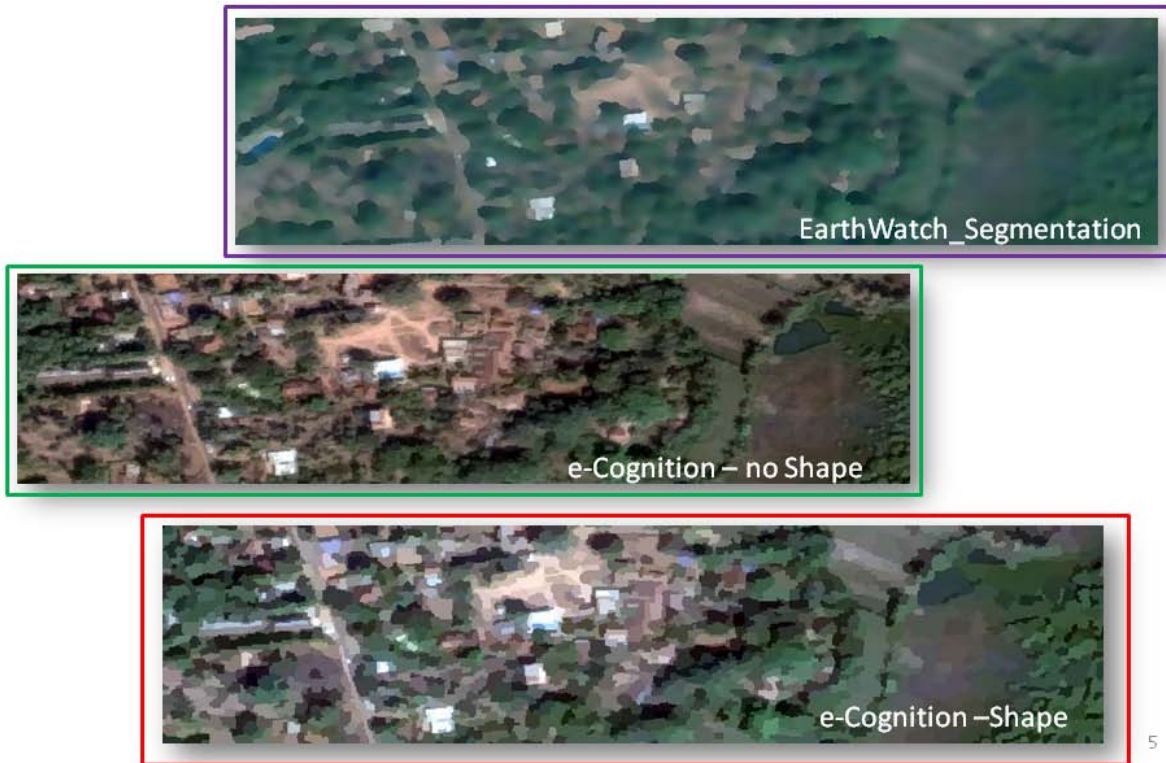
6.1.4 Ο.Α.Κ.Α. Ολυμπιακές Εγκαταστάσεις, Αθήνα



Εικόνα 6.4 Σύγκριση του Αλγορίθμου EarthWatch-MeanShift με δύο εναλλακτικές παραμέτρους στο eCognition

- Αστικό περιβάλλον χωρίς μικρά αντικείμενα προς αναγνώριση μέσω της κατάτμησης, συνεπώς ο αλγόριθμος αναμένεται να λειτουργεί ικανοποιητικά. Το αποτέλεσμα επιβεβαιώνει την αρχική θεώρηση, για τις βασικές οντότητες της εικόνας (στάδιο, χορτάρι, αυτοκίνητα κτλ.) ενώ το eCognition, στην προσπάθεια να εξάγει περισσότερη πληροφορία, δημιουργεί κατατμήσεις που δεν ήταν ιδιαίτερα χρήσιμες, σε μια ενδεχόμενη κατηγοριοποίηση.

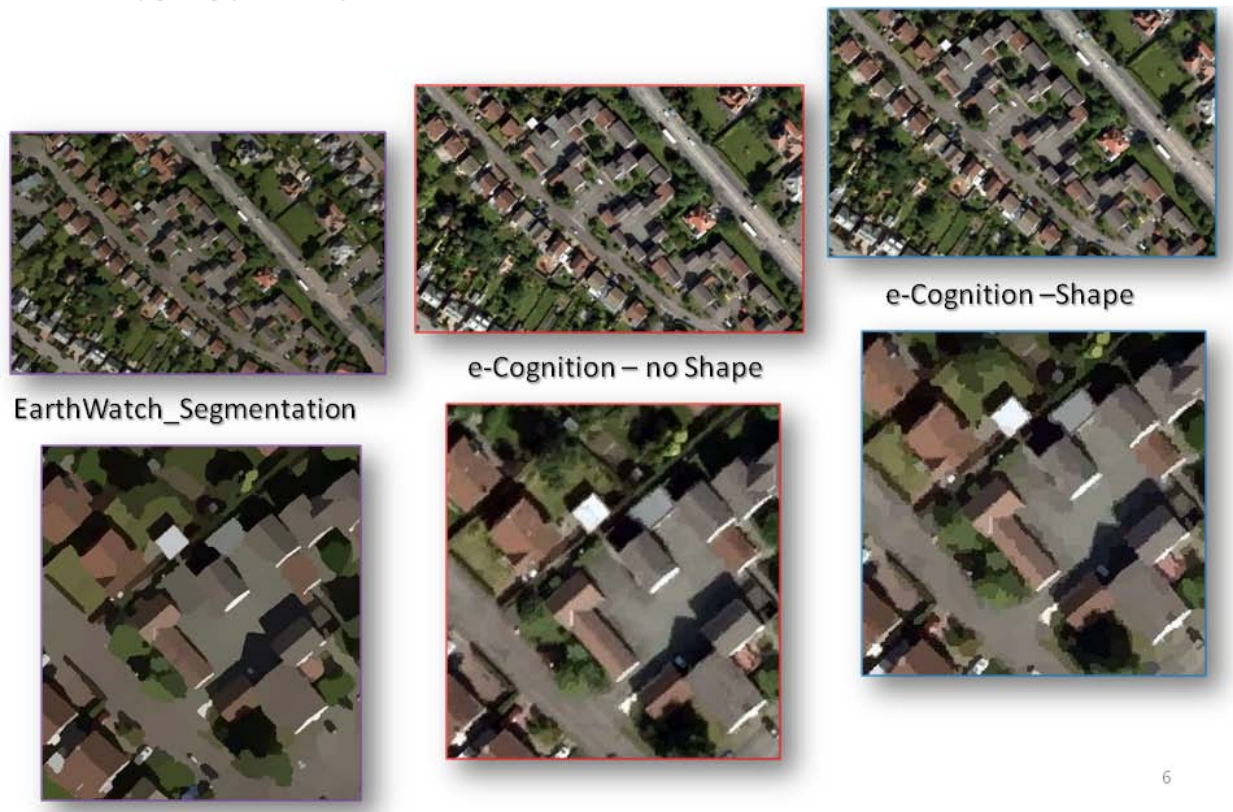
6.1.5 Γοα, Δυτική Ινδία



Εικόνα 6.5 Σύγκριση του Αλγορίθμου EarthWatch-MeanShift με δύο εναλλακτικές παραμέτρους στο eCognition

-
- Μια δύσκολη εφαρμογή για τον Earthwatch αλγόριθμο, μιας και πρέπει να εντοπίσει συγκεκριμένα σχήματα, χωρίς ιδιαίτερες χρωματικές αποκλίσεις από το περιβάλλον τους. Πράγματι στο αποτέλεσμα είναι ορατό το πλεονέκτημα που έχει το eCognition, χωρίς ωστόσο αυτό να σημαίνει ότι δεν μπορεί να αξιοποιηθεί και σε αυτή την περίπτωση ο αλγόριθμος – κάτω φυσικά από περιορισμένες συνθήκες.

6.1.6 Εδιμβούργο, Ηνωμένο Βασίλειο



Εικόνα 6.6 Σύγκριση του Αλγορίθμου EarthWatch-MeanShift με δύο εναλλακτικές παραμέτρους στο eCognition

- Ιδανική εφαρμογή για τους περισσότερους αλγορίθμους κατάτμησης, είναι προφανές ότι το Earthwatch, όπως και το eCognition, δεν δυσκολεύεται καθόλου να εξάγει τη βασική πληροφορία από την εικόνα.

6.1.7 West Ghetto, Ναϊρόμπι



Εικόνα 6.7 Σύγκριση του Αλγορίθμου EarthWatch-MeanShift με δύο εναλλακτικές παραμέτρους στο eCognition

- Αντίθετα με την προηγούμενη εφαρμογή, το γκέτο στη συγκεκριμένη περίπτωση αποτελεί απαγορευτική εφαρμογή για ένα μονο-επίπεδο αλγόριθμο κατάτμησης. Η απόλυτα συγκεχυμένη πληροφορία, δυσκολεύει αρκετά τόσο τον Earthwatch, όσο και τους eCognition αλγορίθμους στην πραγματοποίηση μιας χρήσιμης κατάτμησης.

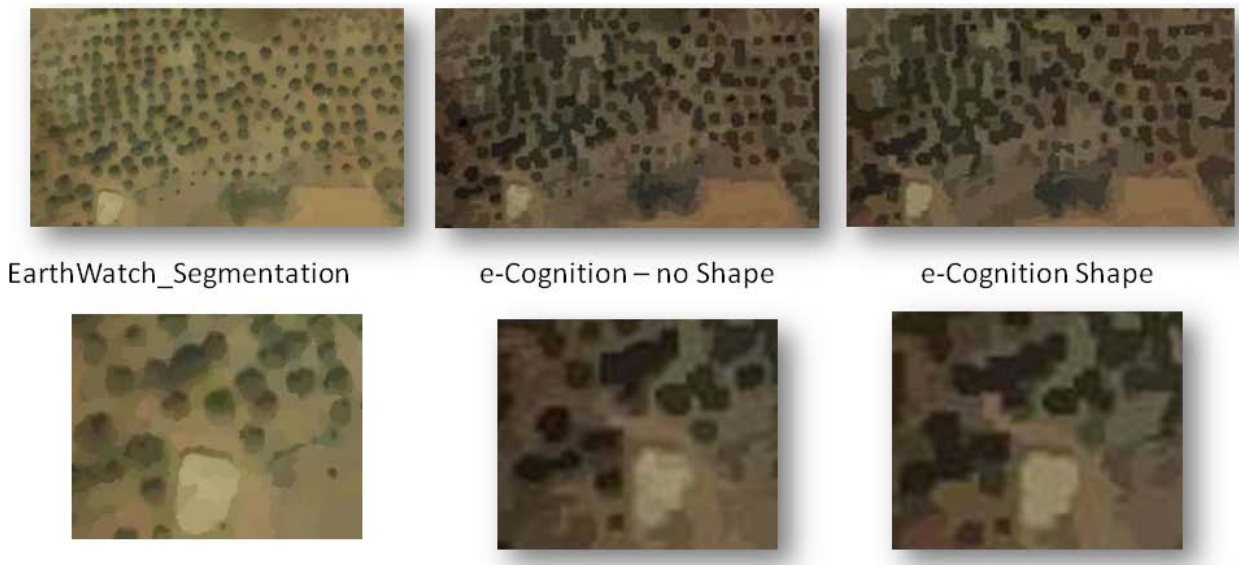
6.1.8 Πόλη της Βενετίας, Ιταλία



Εικόνα 6.8 Σύγκριση του Αλγορίθμου EarthWatch-MeanShift με δύο εναλλακτικές παραμέτρους στο eCognition

- Εφάμιλλη περίπτωση του Εδιμβούργου, με συνδυασμό του ποτάμιου συστήματος, το οποίο ωστόσο δεν αποτέλεσε κανένα απολύτως πρόβλημα. Τα διάφορα πλοία είναι γενικώς ευδιάκριτα και στις τρεις περιπτώσεις, με το Earthwatch ωστόσο να υπολείπεται μικρολεπτομέρειας στο αποτέλεσμα του.

6.1.9 Ηράκλειο, Κρήτη



Εικόνα 6.9 – Σύγκριση του Αλγορίθμου EarthWatch-MeanShift με δύο εναλλακτικές παραμέτρους στο eCognition

-
- Οι ελιές στην τελευταία εφαρμογή αποτέλεσαν πραγματική πρόκληση για τον αλγόριθμο κατάτμησης του Earthwatch. Εξοπλισμένος μόνο με χρωματική ευαισθησία καταφέρνει ένα αποτέλεσμα το οποίο θα μπορούσε να υποβοηθήσει μια κατηγοριοποίηση. Το eCognition από την άλλη, έστω κι αν απαιτεί πολύ μεγάλη διαφοροποίηση στους συντελεστές των κριτηρίων του, δίνει πιο επεκτάσιμα αποτελέσματα.

6.2 Συμπεράσματα για την κατάτμηση εικόνας

6.2.1 Γενικά Συμπεράσματα

Η βασισμένη στο mean shift τεχνική ανάλυσης του χώρου των χαρακτηριστικών είναι ένας αλγόριθμος γενικής χρήσης ο οποίος δεν περιορίζεται μόνο στις, γεωγραφικού περιεχομένου, εφαρμογές που αναφέρθηκαν παραπάνω, αλλά χρησιμοποιείται με επιτυχία και σε άλλες διεργασίες όρασης χαμηλού επιπέδου όπως είναι η κατωφλίωση και η ανίχνευση ακμών, ο εντοπισμός προσώπων (face tracking), η ανάκτηση εικόνας, η κωδικοποίηση βίντεο για MPEG-4, η ανίχνευση και αναγνώριση σχήματος καθώς και η ανάλυση υφής.

Από τη στιγμή που το αποτέλεσμα της εξόδου ελέγχεται μόνο από το εύρος της συνάρτησης πυρήνα, δηλαδή την ακρίβεια της ανάλυσης, η τεχνική μπορεί επίσης να ενσωματωθεί εύκολα σε πολύπλοκα συστήματα όρασης όπου ο έλεγχος παραχωρείται σε μία διεργασία κλειστού βρόχου. Πρόσθετη γνώση για την επιλογή του εύρους μπορεί να αποκομισθεί δοκιμάζοντας τη σταθερότητα της mean shift κατεύθυνσης με διαφορετικά εύρη (Ενότητα 5.6)

Η ολοκληρωμένη λύση για μια αυτόνομη κατάτμηση εικόνας είναι ο συνδυασμός μιας τεχνικής επιλογής εύρους (όπως αυτές που περιγράφηκαν στην ενότητα 2.3.1) με μια ελεγχόμενη και υψηλού-επιπέδου πληροφορία. Σ' αυτήν την περίπτωση, κάθε διαδικασία mean shift σχετίζεται με μια συνάρτηση πυρήνα που «ταιριάζει» καλύτερα στην τοπική δομή του διασυνδεδεμένου πεδίου.

Στην πράξη, ο τελικός στόχος της ανάλυσης είναι άμεσα εξαρτώμενος από το επιθυμητό τελικό αποτέλεσμα, επομένως έχοντας μια προσέγγιση top-bottom, το επιθυμητό έργο θα καθορίσει την παραθυρική ακτίνα που θα χρησιμοποιηθεί στην ανάλυση.

Συμπερασματικά, η διαδικασία mean shift είναι μια πολύτιμη υπολογιστική μονάδα η ευελιξία της οποίας μπορεί να την καταστήσει σημαντικό συστατικό για κάθε εργαλειοθήκη υπολογιστικής όρασης.

6.2.2 Παρατηρήσεις

- Οι χρωματικά εξομαλυμένες περιοχές, οι οποίες διακόπτονται από έντονες περιοχές διαφορετικών χρωματικών διαβαθμίσεων, είναι οι ιδανικές συνθήκες λειτουργίας για τους περισσότερους από τους υπάρχοντες αλγόριθμους. Δεδομένου αυτού οι ημιαστικές περιοχές, βρίσκουν την απόλυτη εφαρμογή στην ανάλυση εικόνας.
- Ο ανθρωπογενής παράγοντας, μετατρέπει και χρησιμοποιεί την κατάτμηση σαν ένα πρόωρο βήμα της αναγνώρισης αντικειμένων, ωστόσο οι υπολογιστικοί αλγόριθμοι κατάτμησης, δεν το κάνουν αυτό. Χρειάζεται ένα είδος έμπειρου συστήματος το οποίο θα δώσει τέτοιες δυνατότητες σε ένα υπολογιστή. Ένα τέτοιο ολοκληρωμένο σύστημα περιγράφεται στο τέλος του κεφαλαίου.
- Για πιο πολύπλοκες εικόνες, οι μηχανές κατάτμησης, προσφέρουν ένα πολύ λιγότερο αξιόπιστο δείκτη για την ανάλυση μεγάλων επιφανειών, και γι αυτό η ωφέλεια τους τίθεται υπο αμφισβήτηση. Η υπεραπλούστευση μιας τέτοιας εικόνας, ή η υπερκατάτμηση, είναι προφανείς και ορατοί κίνδυνοι κατά την ανάλυση. Η αξιολόγηση των αλγορίθμων κατάτμησης, θα πρέπει να γίνεται πάντα με δεδομένη βάση δεδομένων εικόνων.
- Η ποιότητα των κατατμήσεων είναι άμεσα εξαρτώμενη από την ποιότητα της αρχικής εικόνας.
- Γενικά, η χρήση του αλγορίθμου οφείλει να έπεται ενός σταδίου προ-επεξεργασίας της εικόνας, για να επιτυγχάνεται μια πρώτη σχετική ομαλοποίηση των δεδομένων. Όταν ωστόσο δεν υπάρχει διαθέσιμη αρχική πληροφορία η βέλτιστη παραθυρική ακτίνα μπορεί να επιτευχθεί ως το κέντρο του μεγαλύτερου –σε λειτουργία- φάσματος που έχει το ίδιο αριθμό ομάδων συσσώρευσης για τα δεδομένα [9]. Στην πράξη, ωστόσο, ο τελικός στόχος της ανάλυσης είναι άμεσα εξαρτώμενος από το επιθυμητό τελικό αποτέλεσμα, επομένως έχοντας μια προσέγγιση top-bottom, το επιθυμητό έργο θα καθορίσει την παραθυρική ακτίνα που θα χρησιμοποιηθεί στην ανάλυση.

6.3 Συνολικά

Ο αλγόριθμος Mean Shift, αποτελεί ένα ισχυρό εργαλείο στην προσπάθεια κατάτμησης μιας εικόνας.. Έχει τη δυνατότητα να αναζητήσει μέσα από μια τυχαία κατανομή σημείων των δεδομένων με τη βοήθεια του στατιστικού εκτιμητή πυκνότητας, καταλήγοντας στο σημείο όπου η διανομή αυτών είναι πιο πυκνή. Αυτό που θα έπρεπε να τονιστεί σαν εμπειρία από τα διάφορα πειράματα που περιλαμβάνονται στην παρούσα εργασία, είναι το γεγονός ότι η εξαγωγή συγκεκριμένων χαρακτηριστικών από μια εικόνα, είναι αποτέλεσμα της χαρτογράφησης της χρωματικής έντασης των εικονοστοιχείων με βήμα συγκεκριμένου εύρους. Είναι προφανές ότι μια εικόνα έχει πολλές δυνατότητες και απαιτεί πολλές μετατροπές για να καταφέρει κάποιος να χαρτογραφήσει όλα τα επιθυμητά αντικείμενα. Ανεξάρτητα όμως από το είδος του χώρου που

καταλαμβάνει το αντικείμενο, που πρόκειται να εξαχθεί, για όσο διάστημα ο αλγόριθμος δύναται να εντοπίσει τις πυκνοκατοικημένες περιοχές στο εσωτερικό αυτού του χώρου, η επαναληπτική αναζήτηση θα εξαγάγει τελικά το αντικείμενο. Αυτό έχει σημαντικές επιπτώσεις στην κατάτμηση εικόνας, γιατί το μόνο που χρειάζεται είναι να εφαρμοστεί μια τεχνική για όλους τους χαρακτηριστικούς χώρους. Στην παρούσα εφαρμογή δίνεται έμφαση σε πειράματα πάνω σε δεδομένα τηλεπισκόπησης, δίνοντας μια πρώτη εικόνα για τις δυνατότητες του αλγορίθμου. Αυτό σε συνδυασμό με την ευκολία χρήσης του αλγορίθμου πάνω στο περιβάλλον Google Earth, δίνουν μια εναλλακτική προσέγγιση στην αναγνώριση των χαρακτηριστικών μιας δορυφορικής εικόνας.

Η ικανότητα του αλγορίθμου να «επιβιώσει» σε τυχαίας κατανομής φύσεως δεδομένα, εφαρμόζοντας την αναζήτηση μέσης μετατόπισης, προστατεύει την ανάλυση της εικόνας από τον θόρυβο, αυξάνει την προσαρμοστικότητα αυτής, και επιβάλλει μια σχετικά ισχυρή ταυτοποίηση των στοιχείων/τμημάτων της εικόνας. Σαν αντίβαρο της ικανότητας αυτής, είναι ότι ο κύριος όγκος της δυσκολίας έγκειται στη μετάφραση των ελεγχόμενων σημασιολογικών περιορισμών οι οποίοι είναι συνήθως ασαφείς και ανακριβείς. Αυτό αυξάνει το σφάλμα στην μαθηματική έκφραση οι οποίες μπορούν να ενσωματωθούν στη βελτιστοποίηση των διαδικασιών κατάτμησης.

6.4 Μελλοντικές Εργασίες

Οι σκέψεις για επέκταση της συγκεκριμένης εργασίας, οργανώνονται σε δύο κεντρικούς άξονες. Ο ένας αφορά την ανάπτυξη του αλγορίθμου, συμπεριλαμβανομένης της βελτίωσης απόδοσης, δυνατοτήτων και παραμετροποιήσεων, ενώ ο δεύτερος αφορά την διαλειτουργικότητα αλλά και ολοκλήρωσης του συστήματος.

6.4.1 Βελτίωση του Αλγορίθμου Κατάτμησης.

Μέσα από τις διάφορες εφαρμογές, επαληθεύτηκε η θεωρία η οποία θέτει τον mean shift αλγόριθμο σαν ένα από τους πιο αποτελεσματικούς τρόπους προσέγγισης της κατάτμησης δεδομένων. Τα περιθώρια βελτίωσης ωστόσο είναι πολλά. Η δυνατότητα ανάλυσης εικόνας σε n -διάστατο χώρο, ο ορισμός των επιθυμητών κλάστερ, ο τμηματικός στατιστικός έλεγχος των αποτελεσμάτων, η εισαγωγή ασαφούς λογικής ή και άλλων παραμέτρων οι οποίες θα στοχεύουν όχι σε καθολικές, αλλά σε πιο προσωποποιημένες εφαρμογές του αλγορίθμου, είναι κάποια από τα σημεία που επιδέχονται βελτίωση, στο εσωτερικό του αλγορίθμου.

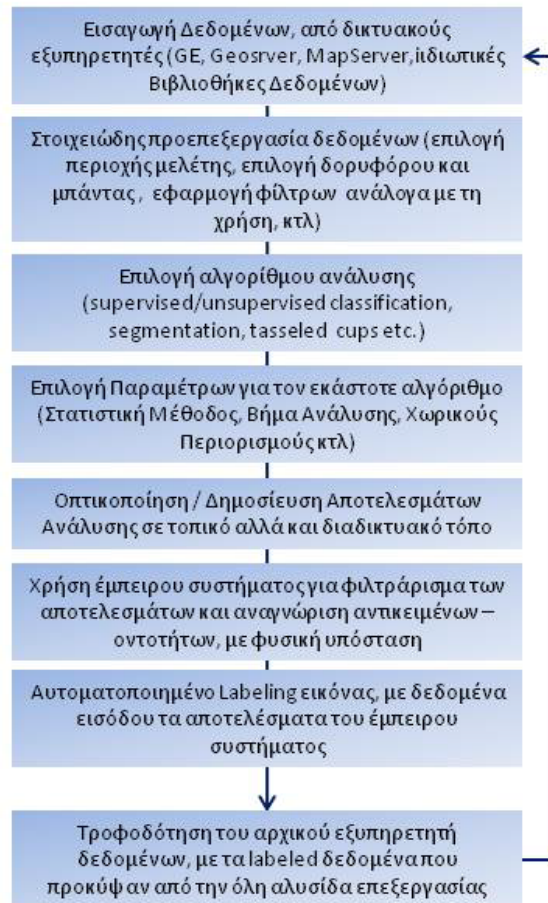
Επίσης, δεδομένου ότι η ανάγκη για κατάτμηση εικόνας, γεννήθηκε από την προσπάθεια ανάπτυξης της computer vision, είναι προφανές ότι μέσα στους κύριους μελλοντικούς στόχους μιας τέτοιας ανάλυσης, είναι να λαμβάνει χώρα σε δεδομένα πραγματικού χρόνου, και όχι σε δεδομένα αρχείου. Για την επίτευξη μιας τέτοιας ιδεατής ανάλυσης, και μάλιστα σε δεδομένα σημαντικού όγκου, και όχι απλές εικόνες των κάποιων Kilobytes, θα πρέπει να αναπτυχθεί η παράλληλη επεξεργασία δεδομένων στο προγραμματιστικό κομμάτι του αλγορίθμου. Είναι προφανές ότι από τη στιγμή που ο χρόνος κατάτμησης μιας εικόνας των 50Mb είναι αρκετά λεπτά, τότε το περιθώριο βελτίωσης είναι ακόμα μεγάλο.

Θεωρητικό και ταυτόχρονα πρακτικό ενδιαφέρον παρουσιάζει η ανάπτυξη και ανάλυση μια προσαρμοστικής διαδικασίας mean shift. Η μεταβαλλόμενη ακτίνα παραθύρου μπορεί να υπολογιστεί βασισμένη σε μια οδηγό συνάρτηση υπολογισμού της πυκνότητας (pilot density estimate).

Μία ακόμα εφαρμογή που διερευνάται είναι η πρόταση για απευθείας τροφοδότηση του νευρωνικού δικτύου του Kohonen με την πληροφορία της πυκνότητας που παρέχει ο αλγόριθμος Mean Shift έτσι ώστε να μη γίνεται μια αυθαίρετη επιλογή των κλάσεων (δηλαδή του τελικού αριθμού των χρωμάτων) αλλά ο αριθμός αυτών να υποδεικνύεται από τα ίδια τα δεδομένα και από τον υπολογισμό της συνάρτησης πυκνότητας πιθανότητας που αυτά σχηματίζουν. Έτσι, ο εντοπισμός των κορυφών θα δείχνει ταυτόχρονα όχι μόνο τον αριθμό των κλάσεων αλλά και τα κέντρα τους, τα οποία θα είναι πλέον γνωστά για το νευρωνικό δίκτυο του Kohonen και δε θα χρειάζεται να προβεί στον υπολογισμό τους μετά από μια αυθαίρετη επιλογή από το χρήστη του αριθμού τους.

6.4.2 Βελτίωση – Ολοκλήρωση του συστήματος ανάλυσης

Η διαλειτουργικότητα διαφορετικών συστημάτων ήταν συνεχώς σε εξέχουσα θέση κατά το σχεδιασμό της μεταπτυχιακής εργασίας. Ένας από τους βασικούς σκοπούς της έμπνευσης και ενασχόλησης με το συγκεκριμένο θέμα, είναι ότι υπάρχουν πολλές οντότητες (συστήματα, λογισμικά, αλγόριθμοι κτλ) οι οποίες κάνουν απομονωμένες τη δουλειά για την οποία δημιουργήθηκαν, χωρίς όμως να υπάρχει κάποια περεταίρω αξιοποίηση των αποτελεσμάτων (πχ ένταξη αυτών σε ένα ευρύτερο σύστημα ανάλυσης). Αν και αυστηρά θεωρητικό, το παρακάτω σύστημα που περιγράφεται, έχει βασιστεί πάνω σε υπάρχουσα βιβλιογραφία.



Ο κύριος όγκος της εργασίας, ασχολείται με τα βήματα 1 έως 5, δηλαδή γίνεται μια συνεργασία συστημάτων, λογισμικών και αλγορίθμων ικανή να εκτελέσει ένα πείραμα/ ανάλυση, και να οπτικοποιήσει τα αποτελέσματα. Στη συνέχεια, τα τρία επόμενα βήματα, θα ολοκληρώσουν το σύστημα επεξεργασίας, μιας και θα αυξήσουν τη χρηστικότητα αυτού κατά πολύ. Πλέον θα αποτελεί μια ολοκληρωμένη λύση πάνω στην ανάλυση εικόνας, της οποίας η βιωσιμότητα κρίνεται δεδομένη, αφενός επειδή θα είναι δωρεάν, αφετέρου, επειδή θα είναι συνεχώς εξελιξιμη, αξιοποιώντας κάθε ένα χρήστη που θα επιλέγει να την χρησιμοποιεί.

6.5 Επίλογος

Η προσπάθεια που έλαβε χώρα σε αυτήν την μεταπτυχιακή εργασία, δείχνει να παίρνει τη θέση της μεταξύ άλλων αντίστοιχων προσπαθειών, ανοικτού λογισμικού, που γίνονται παράλληλα σε όλο τον κόσμο. Λίαν συντόμως, το σπίτι του συγκεκριμένου λογισμικού θα είναι το www.sourceforge.net, ακριβώς όπως και η αρχική έκδοση του προγράμματος, έτσι ώστε ο κάθε ενδιαφερόμενος να μπορεί να κατεβάσει, να χρησιμοποιήσει αλλά και να αναπτύξει τις ιδιότητες του προγράμματος. Με αρχικό σκοπό την διεύρυνση της χρήσης ανώτερων αλγορίθμων ανάλυσης

εικόνας χωρίς την απαραίτητη χρήση δαπανηρών λογισμικών, το προϊόν της εργασίας μπορεί να χρησιμοποιηθεί σε μικρής ή μεσαίας κλίμακας εφαρμογές, μπορεί να εξάγει χρήσιμη πληροφορία, και πάνω από όλα, μπορεί να εξελιχθεί, να μετατραπεί και να παραμετροποιηθεί σχετικά εύκολα.

7 Βιβλιογραφία

- [CHEN95] Cheng, Y., “Mean Shift, Mode Seeking, and Clustering”, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 17, No. 8, pp. 790-799, Aug. 1995.
- [COME97] Comaniciu, D., Meer, P., “Robust Analysis of Feature Spaces: Color Image Segmentation”, *Proc. 1997 Conf. Computer Vision and Pattern Recognition*, pp. 750-755, June 1997.
- [COME99] Comaniciu, D., Meer, P., “Distribution Free Decomposition of Multivariate Data”, *Pattern Analysis and Applications*, Vol. 2, pp. 22-30, 1999.
- [COMM99] Comaniciu, D., Meer, P., “Mean Shift Analysis and Applications”, *Proc. Seventh Int’l Conf. Computer Vision*, pp. 1197-1203, Sept. 1999.
- [COME02] Comaniciu, D., Meer, P., “Mean Shift: A Robust Approach Toward Feature Space Analysis”, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 24, No. 5, pp. 603-618, May 2002.
- [DXTX04] Duan, L., Xu, M., Tian, Qi, Xu, C., “Mean Shift Based Nonparametric Motion Characterization”, *Intl Conf Image Processing*, Vol. 3, pp. 1597-1600, 2004.
- [CHLA04] Chow, T., Lam, K., “Mean-Shift Based Mixture Model for Face Detection in Color Image”, *Intl Conf Image Processing*, Vol. 1, pp. 601-604, 2004
- [SKOT05] Σκότιδα, Χ. «Αλγόριθμος Meanshift Και Η Εφαρμογή Του Στη Μείωση Χρωμάτων Σε Εικόνες». σσ. 20 - 46

7.1 Διαδικτυακοί Χώροι

7.1.1 Γενικού Περιεχομένου

[An Interactive Java Statistical Image Segmentation System: Gemident](#) [MATLAB Central - Files matlab image segmentation -](#) [MATLAB Central - File detail - Image segmentation & Extraction](#) [MATLAB Central - Files](#) [MATLAB Central - File detail - seed based segmentation](#) [MATLAB Central - File detail - MRI Brain Segmentation](#) [MATLAB Central - File detail - image segmentation using iterative watershed plus ridge detection](#) [MATLAB Central - File detail - Hill-Climbing Color Image Segmentation](#) [MATLAB Central - File detail - Active Contour Segmentation](#) [MATLAB Central - File detail - Wavelet based image compression using segmentation](#) [MATLAB Central - File detail - level set for image segmentation](#) [MATLAB Central - File detail - Image Segmentation via topological derivative](#) [MATLAB Central - File detail - Edge Detection and Segmentation](#) [Mean Shift Segmentation in Matlab | Shawn Lankton Online](#) [edison image segmentation - Αναζήτηση Google](#) [Edge Detection and Image Segmentation \(EDISON\) System](#) [Edge Detection and Image Segmentation \(EDISON\) System](#) [Edge Detection and Image Segmentation \(EDISON\) System](#) [MATLAB Central - File detail - Image segmentation & Extraction](#)

7.1.2 Αλγόριθμος Μέσης Μετατόπισης

[Mean Shift Segmentation in Matlab | Shawn Lankton Online](#) [Mean Shift Filter](#) [Mean Shift Filter article](#) [SelectSingle science](#) [ScienceDirect - Pattern Recognition Letters : Unsupervised color-texture segmentation based on soft criterion with adaptive mean-shift clustering](#) [CiteULike: Segmentation of Color](#)

[Images using Mean Shift Algorithm for Feature Extraction](#) [Mean Shift: A Robust Approach toward Feature Space Analysis](#) [articleSelectSingle science](#) [articleSelectSingle science](#) [ScienceDirect - Computers and Electronics in Agriculture : Mean-shift-based color segmentation of images containing green vegetation](#) [mean shift track.ppt](#) [MeanShift.ppt](#) [Mean Shift Filter](#) [MSB: A mean-shift-based approach for the analysis of structural variation in the genome — Genome Research](#) [Mean Shift Based Image Segmenter](#) [Mean Shift Segmentation in Matlab | Shawn Lankton Online](#) [Gaussian Mean-Shift Is an EM Algorithm](#) [mean shift software - Αναζήτηση Google](#) [implement mean-shift clustering algorithm in java - Software Development Jobs - oDesk](#) [A New Improvement on Mean-ShiftAlgorithm](#) http://clef-campaign.org/2006/working_notes/workingnotes2006/maillotCLEF2006.pdf - [Με την υποστήριξη των Εγγράφων Google](#)

7.1.3 **Stitching images java source**

[automatic image stitching algorithm open source - Αναζήτηση Google](#)

[How to Stitch Photos Using Free or Open-Source Software - KAPwiki](#)

[PT Open Gui](#)

[Browse hugin Files on SourceForge.net](#)

[PTopenGUI screenshots](#)

[quanjingpinjie a robust automatic image stitching algorithm integration of a fully so Graph program](#)
www.pudn.com

[login_do.asp](#)

[Upload source code/document - www.pudn.com](#)

[java source stitching image - Αναζήτηση Google](#)

[Image Stitching - with Source](#)

[Image stitching > Make Montage?: msg#00093 java.imagej](#)

[Features](#)

[Plugins](#)

[Stitch BLISS Tile Images](#)

[Homepage of Stephan Preibisch at MPI-CBG](#)

7.1.4 GDal_wrapping_utility

[GDAL: GDALWarpOptions Struct Reference](#)

[GDAL: GDALWarpOptions Struct Reference](#)

[GDAL: GDALWarpOperation Class Reference](#)

[gdalwarp - Αναζήτηση Google](#)

[GDAL: GDAL Warp API Tutorial](#)

[GDAL: GDAL Data Model](#)

[GDAL: GDALWarpOptions Struct Reference](#)

[GDALAutoCreateWarpedVRT - Αναζήτηση Google](#)

[GDAL: gdalwarper.h File Reference](#)

[gdal warp image height - Αναζήτηση Google](#)

[Ubuntu Manpage: gdalwarp - gdalwarp simple image reprojection and warping utility](#)

[Bending the Earth — gdalwarp and the Blue Marble](#)

[Global Maps](#)

[Using gdalwarp](#)

[PROJ.4 - Trac](#)

[FAQ](#)

[GeodesicCalculations - PROJ.4 - Trac](#)

[Welcome - GeoServer](#)

[calculator.html](#)

<http://software.informer.com/manage/register.php>

[CodeProject: Embedding Google Earth in a C# Application. Free source code and programming help](#)

[Open Source Geospatial Software in Java](#)

[UltraEdit scripting engine tutorial](#)

[Beginner's guide to writing and building your first C/C++ application in UESTudio IDE](#)

7.1.5 Edison_Source_References

[edison image segmentation - Αναζήτηση Google](#)

[Edge Detection and Image Segmentation \(EDISON\) System](#)

[EDISON Code](#)

[Edge Detection and Image Segmentation \(EDISON\) System](#)

[Edge Detection and Image Segmentation \(EDISON\) System](#)

[Links](#)

[Robust Analysis of Visual Data](#)

[Links](#)

[Links](#)

[The Computer Vision Homepage](#)

[RIUL, CAIP, RUTGERS](#)

[Ph.D. Thesis DORIN COMANICIU](#)

[Theses](#)

[Tutorials](#)

[rotechcv.pdf \(application/pdf Object\)](#)

[tutorialica.pdf \(application/pdf Object\)](#)

[tutorialrpca.pdf \(application/pdf Object\)](#)

7.1.6 GE

[Google Earth Hacks - File Downloads - Most Popular files](#)

[A Face in the Clay - Top 10 Google Earth Finds - TIME](#)

[UFO Landing Pads, Maybe? - Top 10 Google Earth Finds - TIME](#)

[Oprah Maze - Top 10 Google Earth Finds - TIME](#)

[50.010489, -110.116906 - Google Maps](#)

[WEKA Classification Algorithms](#)

8 Παράρτημα I

8.1 Earth Watch - Design and Implementation

8.1.1 Overall Design and Implementation

Earth Watch is designed as a standalone windowed application with emphasis taken to integrate it with Google Earth as much as possible to act like a native plug-in. It was designed to fit and work around the windowing system used in Google Earth, by docking to the left side of the screen and not being so wide that it would encroach on the Google Earth render window. As such it is a standalone program that will load Google Earth if need be at runtime.

The interface provided by Google Earth, is only for their Windows client, and consists of a set of COM routines primarily for navigation. Support for interaction with “features” such as image overlays, polygonal overlays, and access to screen shots is not fully implemented. As a result, a lot of tricks had to be employed to achieve the level of interfacing to make this application possible. For one thing, while there are COM routines for taking screen shots, the result degrades if used too often and is not high resolution. Another is that, access to “features” is limited to their enumeration, enablement and disablement. The work around for the screen shots was to get the HWND reference, which the COM interface provided, and use Win32 GDI to get access to the display buffer.

The work around for access to features proved to be more complicated. Read access was accomplished by accessing the KML data file that Google Earth stores. The trick was to simulate the keystroke Ctrl-Shift-S which ensures that the features are saved to the data file. Write access was achieved by first sending the KML data to Google Earth through COM, which stored the new features in a “Temporary Places” folder. So that the features would be accessible later on by a subsequent read operation, they needed to be escalated from the Temporary Places folder to the “My Places” folder. This was accomplished by simulating a sequence of simulated keystrokes.

2 OO Model

Strict Object Oriented principals were adhered to in the development of Earth Watch with data encapsulation and where possible the code was modularized into class units. Additionally a strict naming convention was utilized. This naming convention included using the scope as well as the type along with the name of the variable. Now a brief overview of the modules within Project Earth Watch and a requisite subproject the Google Earth Interface.

8.1.2 Data Model

The data model for Earth Watch starts with a folder within the My Documents folder of the user called “EarthWatchExperiments”. Within this folder, each experiment is given its own folder after its name. Within each experiment folder the following files exist:

<experiment name>.DAT - This file contains the list of classes within the experiment, the experiment region, the polygons used for training the classes, and the resultant evaluation of the classes over the experiment region.

<experiment name>.ARFF – This file is the stochastically drawn data set used to build the classifier within Weka.

<experiment name>.BMP – This is the before picture taken over the experiment region used within the PDF report that is generated.

<experiment name>-overlay.BMP – This is the color coded by class image overlay.

<experiment name>.KML– This file represents the objects rendered within Google Earth to represent this experiment.

<experiment name>.PDF– This is the report generated by the experiment illustrating the before and after classification images, the accuracy of the trainer, and the relative proportions of the classes within the experiment region.

Survey/Survey.DAT – This file is the database file containing the information pertaining to the conducted experiment. It stores the resolution that the survey was conducted at, the resolution and number of images used to cover the whole area at the desired resolution, and the bounding boxes representing the geographic regions that each image covers.

Survey/SurveyImage-X_Y.BMP – These are the picture files for the Survey.

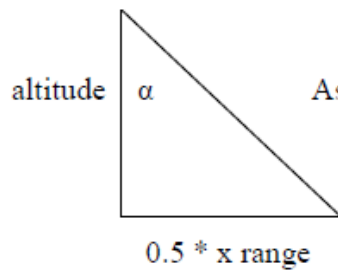
8.1.3 Navigation

With regard to navigation there are two issues at hand within the scope of this project. The first being the graphics pipeline used to project the Earth on to a 2-d display. The second being the model used to represent the surface of the earth.

With respect to the second issue the issue is over coordinate system. With Latitude/Longitude/Altitude, the system with which coordinates are represented in Google Earth, a 1 degree offset in Longitude does not directly correlate to a corresponding constant offset in meters to the right/left. Instead this is dependent on the Latitude, where as the point in question approaches the poles, the difference in meters between two points of the same Latitude but with a 1 degree offset in Longitude approaches zero. So the question becomes: How does one translate or adequately calculate the distance between two points on the surface of a sphere or, slightly more accurately on the surface of an ellipsoid.

At first the Universal Transverse Mercator system (UTM) was utilized. This provided the feature of translating Longitude and Latitude into an Easting and a Northing where the Northing represents the distance from the Equator in meters and the Easting represents the Distance from an arbitrary point depending on which of the 60 zone you happened to be in. With the ability to calculate now the distance between two points with meters as a unit (instead of a difference in latitude and longitude), something that would for multiple places in the globe, I was provided with enough power to determine the aperture angle that was used to project Google Earth onto a 2 dimensional screen experimentally.

Simple trigonometry illustrates the relationship between the altitude and the viewable area in terms of this angle.



As you can see here, the aperture angle:

$$2\alpha = 2\tan^{-1}(0.5x\text{-range}/\text{altitude})$$

(With numerous experiments at different places on the globe feeding in altitudes and observing the x-range, α was experimentally calculated to be 30 degrees)

There is of course another aperture angle on the y-range but as it turned out it was directly related to the aspect ratio of the device.

After that UTM was abandoned for fear of having to navigate across zonal boundaries. While there is some overlap with zonal boundaries, done so navigators who had to navigate across zone boundary could pick one zone in which to do their calculations, this was not done in the spirit of the potentially large distances that I would be navigating, but more for ships and vessels calculating their next bearing. Calculations for the arc-length of an ellipsoid were considered but ultimately I implemented the tried and true haversine function which proved to be accurate enough for the navigation purposes of this project. 15% overlap in survey images more than covers the margin of error of the spherical approximation.

WEKA_2 algorithm – Reference

8.2 Learning Vector Quantization (LVQ)

LVQ algorithm was the first algorithm implemented for the WEKA platform. This section contains some notes regarding the implementation of the LVQ algorithm in WEKA, taken from the initial release of the plug-in (back in 2002-2003). LVQ Weka

8.2.1 Learning Vector Quantization

- A competitive learning algorithm said to be a supervised version of the Self-Organizing Map (SOM) algorithm by Kohonen

- Goal of the algorithm is to approximate the distribution of a class using a reduced number of codebook vectors where the algorithm seeks to minimise classification errors
- Codebook vectors become exemplars for a particular class - attempting to represent class boundaries
- The algorithm does not construct a topographical ordering of the dataset (there is no concept of explicit neighborhood in LVQ as there is in the SOM algorithm)
- Algorithm was proposed by Kohonen in 1986 as an improvement over Labelled Vector Quantization
- The algorithm is associated with the neural network class of learning algorithms, though works significantly differently compared to conventional feed-forward networks like Back Propagation

8.2.2 Advantages of the Learning Vector Quantization algorithm

- The model is trained significantly faster than other neural network techniques like Back Propagation
- It is able to summarise or reduce large datasets to a smaller number of codebook vectors suitable for classification or visualisation
- Able to generalise features in the dataset providing a level of robustness
- Can approximate just about any classification problem as long as the attributes can be compared using a meaningful distance measure
- Not limited in the number of dimensions in the codebook vectors like nearest neighbour techniques
- Normalisation of input data is not required (normalised may improve accuracy if attribute values vary greatly)
- Can handle data with missing values
- The generated model can be updated incrementally

8.2.3 Disadvantages of the Learning Vector Quantization algorithm

- Need to be able to generate useful distance measures for all attributes (Euclidean is usually used for numeric attributes)
- Model accuracy is highly dependent on the initialisation of the model as well as the learning parameters used (learning rate, training iterations, etcetera)
- Accuracy is also dependent on the class distribution in the training dataset, a good distribution of samples is needed to construct useful models
- It is difficult to determine a good number of codebook vectors for a given problem

8.2.4 Algorithm Notes

Supports 7 implementations of the LVQ algorithm

- **LVQ1** - A single BMU (best matching unit) is selected and moved closer or further away from each data vector, per iteration.
- **OLVQ1 (Optimised LVQ1)** - The same as LVQ1, except that each codebook vector has its own learning rate
- **LVQ2.1** - Two BMU's are selected and only updated if one belongs to the desired class and one does not, and the distance ratio is within a defined window
- **LVQ3** - The same as LVQ2.1 except if both BMU's are of the correct class, they are updated but adjusted using an epsilon value (adjusted learning rate instead of the global learning rate)
- **OLVQ3 (Optimised LVQ3)** - The same as LVQ3 except each codebook vector has its own learning rate in the same manner as OLVQ1
- **Multi-Pass LVQ**- This is the recommended usage of the algorithm (Kohonen) where a quick rough pass is made on the model using OLVQ1, then a long fine tuning pass is made on the model with any of LVQ1, LVQ2.1 or LVQ3
- **Hierarchical LVQ**- This is an idea I had (let me know if it's not original) where an LVQ model is constructed and each codebook vector is treated as a cluster centroid. All codebook vectors are evaluated and a number are selected as candidates for sub-models. Sub models are constructed for all candidate codebook vectors and those sub models that out perform (in terms of classification accuracy) their parent codebook vector are kept as part of the model. During testing a data instance is first mapped onto its BMU, if that BMU has a sub-model (that was not pruned during training), the sub model is used for classification, otherwise the class value in the BMU is used for classification. This algorithm has proven most useful for large datasets where a low-complexity model is used as the base model, and specialised LVQ models are used at each codebook vector to better model the data. Supports any Weka algorithm as the BMU's sub model, not just LVQ.

Supports 2 implementations of the Self-Organizing Map (SOM) algorithm

The Self-Organizing Map (SOM) algorithm is not a classification algorithm, though it can be used for classification tasks. An implementation of the unsupervised SOM algorithm is provided that can apply labels to the map so that it can be used for classification. The SOM implementation also supports supervised learning known as LVQ-SOM where codebook vectors in the neighbourhood of the best matching unit (BMU) that do not match the class of the data instance are pushed further away rather than closer to the data instance. No map visualisation techniques have been provided, but the vectors can be retrieved (debug mode in the GUI, or API call) and displayed using Kohonen's SOM_PAK, in MathLab or your visualisation application of choice.

- **SOM** - Self-Organizing Map algorithm that supports supervised and unsupervised learning and dynamical labelling or post-training map labelling.
- **Multi-Pass SOM**- The recommended usage of the SOM algorithm where two passes are performed on the same underlying model. The first pass is a rough ordering pass with large neighbourhood, learning rate and small training time. The second pass is the fine tuning pass that has a longer training time, small initial neighbourhood value and smaller initial learning rate.

Supports 4 Feed-Forward Neural Network algorithms

The default MultilayerPerceptron implementation provided with Weka is good, but did not meet my needs. I have added feed-forward neural network algorithms that provide a clean implementation, a simpler interface and more readable/maintainable source code. I believe that the back propagation implementation is also faster, though requires explicit specification of neurons at each layer.

- **Perceptron** - Perceptron learning rule, only supports binary data, provides a binary output and uses the Sign transfer function
- **Widrow Hoff** - Widrow Hoff learning rule, only supports binary data, provides a binary output and uses the Step transfer function
- **Back Propagation** - Standard implementation with support for up to 3 layers (user defined), momentum, weight decay, 5 transfer functions, etcetera
- **Bold Driver Back Propagation**- (Vogl's Method) Same as back propagation except the learning rate is dynamically determined

Supports 1 new Filter

NormaliseMidpointZero - Located under UnsupervisedFilters this filter was implemented to scale data values to between -1.0 and +1.0 for use with the hyperbolic tangent (tanh) transfer function in feed forward neural networks

Algorithm Usage Notes and Heuristics

Enable algorithm debugging to see how many of the codebook vectors are actually used by the end of training. This is most useful for knowing when to increase or decrease the number of codebook vectors thus increasing model accuracy.

- Less codebook vectors typically need less training; obviously more codebook vectors need more training iterations.
- Normalisation or standardisation of the dataset may improve the accuracy of an LVQ model, especially those datasets with many (> 15) numeric attributes

- It is usually a good idea to have enough training iterations to expose all training data to the model, usually a few times
- Performing feature selection (removal of unnecessary attributes) before running the algorithm will increase speed
- Interestingly, the SOM algorithm can pick up on features that the LVQ cannot pick up as easily, sometimes producing superior models
- LVQ only supports classification (not regression) type problems
- LVQ supports training and test data with missing values
- The LVQ algorithm can be used as a data reduction technique, where the resulting codebook vectors are taken and used in visualisation techniques such as Sammon's Mapping or a PCA Projection
- Multi-Pass LVQ is the recommended usage of the algorithm (Kohonen), if individual algorithms are used by themselves, longer training times and larger learning rates should be used
- Hierarchical LVQ can be used to over-learn the training data, providing superior accuracy for many datasets. It is possible for this algorithm to learning the training data exactly by setting both percentage parameters to 0 and selecting a base model and sub model that perform well
- When using the feed forward neural network algorithms it is imperative that the dataset are scaled to the range of the selected transfer function. Example; for the Sigmoid transfer function data should be normalised (0 to 1), For Tanh use the new filter NormaliseMidpointZero (-1 to +1).

Algorithm	Parameter	Recommendations
Multi-Pass LVQ		<ul style="list-style-type: none"> • Pass 1 <ul style="list-style-type: none"> ○ OLQV1 recommended for fast approximation ○ intended to quickly and roughly approximate the class distribution ○ larger learning rates and shorter training times should be used ○ learning rate (0.1 to 0.4 - say 0.3) ○ training iterations (30 to 50 times the number of codebook vectors) • Pass 2 <ul style="list-style-type: none"> ○ LVQ1 or LVQ3 recommended for fine tuning ○ intended to slowly and precisely finetune the codebook vectors to the class distribution in the data ○ smaller learning rates and longer training times should be used ○ learning rate (< 0.1 - say 0.05) ○ training iterations (about 10 times that of the first pass)

General

- The number of codebook vectors should be sufficient to summarise the class distribution in the problem (20 to 50 is suitable for most "toy" classification problems, Kohonen uses thousands in some case studies)
- The more complex the class distribution, the more codebook vectors that will be required to approximate the problem
- Good values for the window width on LVQ2.1 and LVQ3 are between 0.2 and 0.3
- The epsilon value in LVQ3 and OLVQ3 should be small for small window values (0.1 to 0.3)
- For many classification problems the multi-pass approach can perform as well as or outperform any of the other single pass LVQ algorithms in half the training time, and with half as many neurons (see example results), thus is recommended for improved results
- OLVQ1 is faster at approximating a solution and should be used as a rough gage of the algorithms performance for a given problem

8.3 Programming Modules and technologies included in this project

8.3.1 Google Earth COM API

This document describes the features of the [Google Earth](#) COM API. The Google Earth COM API allows third party applications to query information from and send commands to Google Earth. Through [ApplicationGE](#), applications can query the current viewport, control the 3D viewpoint, use KML features, and more.

8.3.2 gekmlib

A Java library to ease creation and manipulation of KML files. A KML file can be loaded as simply as this: `Kml kmlRoot = new KMLParser().parse(new File(filename));` Alternatively a Kml document can be built up programmatically, eg by creating new `Placemark()`, `new Folder()` etc and adding them to the root, or other Folders.

`Kml.toKML()` produces the KML document.

`KML.toUpdateKML()` will produce a KML document representing the updates since `toUpdateKML()` was last called.

8.3.3 Stitching of the Images

ImageJ is a public domain Java image processing program inspired by [NIH Image](#) for the Macintosh. It runs, either as an online applet or as a downloadable application, on any computer with a Java 1.4

or later virtual machine. Downloadable distributions are [available](#) for Windows, Mac OS, Mac OS X and Linux.

It can display, edit, analyze, process, save and print 8-bit, 16-bit and 32-bit images. It can read many image formats including TIFF, GIF, JPEG, BMP, DICOM, FITS and "raw". It supports "stacks", a series of images that share a single window. It is multithreaded, so time-consuming operations such as image file reading can be performed in parallel with other operations.

It can calculate area and pixel value statistics of user-defined selections. It can measure distances and angles. It can create density histograms and line profile plots. It supports standard image processing functions such as contrast manipulation, sharpening, smoothing, edge detection and median filtering.

It does geometric transformations such as scaling, rotation and flips. Image can be zoomed up to 32:1 and down to 1:32. All analysis and processing functions are available at any magnification factor. The program supports any number of windows (images) simultaneously, limited only by available memory.

Spatial calibration is available to provide real world dimensional measurements in units such as millimeters. Density or gray scale calibration is also available.

ImageJ was designed with an open architecture that provides extensibility via Java plugins. Custom acquisition, analysis and processing plugins can be developed using ImageJ's built in editor and Java compiler. User-written plugins make it possible to solve almost any image processing or analysis problem.

ImageJ is being developed on [Mac OS X](#) using its built in editor and Java compiler, plus the [BBEdit](#) editor and the [Ant](#) build tool. The source code is freely available. The author, Wayne Rasband (wayne@codon.nih.gov), is at the Research Services Branch, National Institute of Mental Health, Bethesda, Maryland, USA.

8.3.4 Win32_COM reference

Τα λειτουργικά συστήματα που υλοποιούν όλο ή τμήμα του μοντέλου του Win32 υποστηρίζουν *προεκτοπιστική πολυνηματική επεξεργασία* (preemptive multithreading). Αυτό σημαίνει ότι ο πυρήνας του λειτουργικού συστήματος θεωρεί τα νήματα ως κύρια μονάδα εκτέλεσης, τα οποία εκτελούνται εκ περιτροπής για ένα καθορισμένο χρονικό διάστημα. Μεταξύ των λειτουργικών συστημάτων αυτών υπάρχουν διαφορές ως προς την εκτέλεση

παλαιότερων (16-bit) εφαρμογών, αλλά εδώ θα ασχοληθούμε με τις 32-bit εφαρμογές για τις οποίες η αντιμετώπιση είναι κοινή.

ΝΗΜΑΤΑ ΚΑΙ ΔΙΕΡΓΑΣΙΕΣ

Το Win32 διαχειρίζεται *διεργασίες* (processes), κάθε μία από τις οποίες μπορεί να αποτελείται από πολλά *νήματα εκτέλεσης* (threads of execution). Μια διεργασία είναι το στιγμιότυπο μίας εκτελούμενης εφαρμογής και για το ΛΣ αποτελείται από:

- Ένα χώρο εικονικών διευθύνσεων μεγέθους 4 GB (διευθύνσεις 32 bit). Από τα 4GB αυτά, ένα τμήμα είναι διαθέσιμο στη διεργασία, ενώ το υπόλοιπο είναι διευθύνσεις του πυρήνα.
- Τμήματα φυσικής μνήμης αντιστοιχισμένα σε εικονικές διευθύνσεις της διεργασίας.
- Τμήματα του swap file αντιστοιχισμένα σε εικονικές διευθύνσεις της διεργασίας.
- Τμήματα του εκτελέσιμου αρχείου που εκτελεί η διεργασία αντιστοιχισμένα σε εικονικές διευθύνσεις της διεργασίας.
- Δεσμευμένους πόρους του συστήματος όπως ανοιχτά αρχεία, αντικείμενα συγχρονισμού κλπ.
- Τουλάχιστον ένα νήμα εκτέλεσης.

Η βασική μονάδα εκτέλεσης στο Win32 είναι το *νήμα* (thread). Κάθε νήμα «περιέχεται» σε μία διεργασία και έχει πρόσβαση στο χώρο μνήμης αυτής. Εκτός από τη διεργασία στην οποία αντιστοιχεί, το ΛΣ χρειάζεται να γνωρίζει για κάθε νήμα:

- Τον κώδικα που εκτελεί, δηλαδή ουσιαστικά την τρέχουσα τιμή του PC για το νήμα.
- Τη στοίβα εκτέλεσής του.

Τα νήματα είναι οι βασικές οντότητες στις οποίες το λειτουργικό σύστημα αποδίδει χρόνο ΚΜΕ. Για κάθε νήμα το ΛΣ διατηρεί ένα σύνολο δεδομένων όπου κρατείται το *περιεχόμενο* του (context) όσο το νήμα περιμένει να δρομολογηθεί για επεξεργασία. Το περιεχόμενο ενός νήματος αποτελείται από τις τιμές των καταχωρητών, τη *στοίβα του πυρήνα* (kernel stack), μια εγγραφή για το *περιβάλλον* (environment) του νήματος, και τη στοίβα του ίδιου του νήματος που βρίσκεται στο χώρο διευθύνσεων της διεργασίας. Όλα τα νήματα μιας διεργασίας μοιράζονται τον εικονικό χώρο διευθύνσεών της και μπορούν να προσπελάσουν τις παγκόσμιες μεταβλητές της όπως και τους πόρους συστήματος που κατέχει η διεργασία. Όταν δημιουργείται μια διεργασία, τα Windows NT αυτόματα δημιουργούν το πρώτο της νήμα, το οποίο λέγεται και το *κύριο* (primary) νήμα. Αυτό μπορεί μετά να δημιουργήσει επιπλέον νήματα της διεργασίας αυτής και εκείνα με τη σειρά τους άλλα. Αν ένα οποιοδήποτε νήμα μιας διεργασίας δεν έχει τερματίσει, η διεργασία δεν πεθαίνει· μια διεργασία τερματίζεται όταν όλα της τα νήματα έχουν τερματίσει. Τα νήματα μίας διεργασίας μπορούν να δημιουργήσουν άλλες διεργασίες, οι οποίες εκτελούνται με διαφορετικό χώρο εικονικών διευθύνσεων από την αρχική. Μεταξύ των διεργασιών δεν υπάρχει σχέση γονέα-παιδιού όπως σε άλλα ΛΣ.

8.3.5 GDI 32 reference

The **Graphics Device Interface** (GDI) is a [Microsoft Windows application programming interface](#) and core operating system component responsible for representing graphical objects and transmitting them to [output devices](#) such as [monitors](#) and [printers](#).

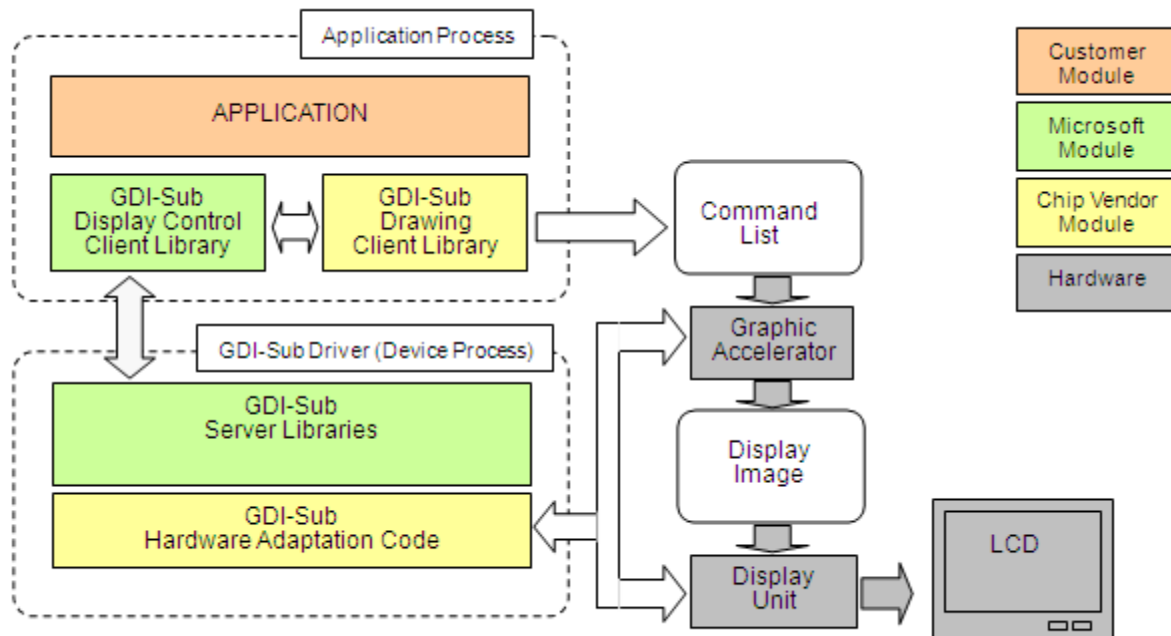
GDI is responsible for tasks such as drawing lines and curves, rendering [fonts](#) and handling [palettes](#). It is not directly responsible for drawing windows, menus, etc.; that task is reserved for the user subsystem, which resides in [user32.dll](#) and is built atop GDI. GDI is similar to Macintosh's [QuickDraw](#).

Perhaps the most significant capability of GDI over more direct methods of accessing the hardware is its scaling capabilities, and abstraction of target devices. Using GDI, it is very easy to draw on multiple devices, such as a screen and a printer, and expect proper reproduction in each case. This capability is at the center of all [What You See Is What You Get](#) applications for Microsoft Windows.

Simple games which do not require fast graphics rendering use GDI. However, GDI cannot animate properly (no notion of synchronizing with the [framebuffer](#)) and lacks [rasterization](#) for 3D. Modern games tend to use [DirectX](#) or [OpenGL](#), which give programmers the capabilities to use features of modern hardware.

Overall Structure of GDI-Sub

GDI-Sub roughly consists of a client library linked with an application and a GDI-Sub driver module, which is loaded as a device driver into device.exe. In addition, GDI-Sub also includes an optional virtual display driver to run a GDI application within the GDI-Sub environment, as well as a software rendering module. Details of these are covered later in this document(refer "11.4.1 GDI-Sub Drawing Client Library", "11.4.2 GDI-Sub Display Control Client Library", "11.4.3 GDI-Sub Server Libraries" and "11.4.4 Hardware Adaptation Code").



Εικόνα: GDI-sub overview diagram

8.3.6 Native Win32C using the Java Native Interface (JNI)

the Java™ Native Interface (JNI). It will be useful to you if you are interested in any of the following:

- integrating a Java application with legacy code written in languages such as C or C++
- incorporating a Java virtual machine implementation into an existing application written in languages such as C or C++
- implementing a Java virtual machine
- understanding the technical issues in language interoperability, in particular how to handle features such as garbage collection and multithreading.

The JNI was initially released in early 1997. The book summarizes two years of collective experience gained by engineers at Sun Microsystems as well as the vast number of developers in the Java technology community. The JNI was designed following a series of discussions between Sun Microsystems and Java technology licensees. The JNI partly evolved from Netscape's Java Runtime Interface (JRI), which was designed by Warren Harris. Many people from Java technology licensee companies actively participated in the design discussions. They include Russ Arun (Microsoft), Patrick Beard (Apple), Simon Nash (IBM), Ken Root (Intel), Ian Ellison-Taylor (Microsoft), and Mike Toutonghi (Microsoft). The JNI design also benefited greatly from Sun internal design reviews conducted by Dave Bowen, James Gosling, Peter Kessler, Tim Lindholm, Mark Reinhold, Derek White, and Frank Yellin. Dave Brown, Dave Connelly, James McIlree, Benjamin Renaud, and Tom Rodriguez made significant contributions to the JNI enhancements in Java 2 SDK

1.2. Carla Schroer’s team of compatibility testers in Novosibirsk, Russia, wrote compatibility tests for the JNI. In the process they uncovered places where the original specification was unclear or incomplete. The JNI technology would not have been developed and deployed without the management support of Dave Bowen, Larry Abrahams, Dick Neiss, Jon Kannegaard, and Alan Baratz.

THE Java™ Native Interface (JNI) is a powerful feature of the Java platform. Applications that use the JNI can incorporate *native code* written in programming languages such as C and C++, as well as code written in the Java programming language. The JNI allows programmers to take advantage of the power of the Java platform, without having to abandon their investments in legacy code. Because the JNI is a part of the Java platform, programmers can address interoperability issues once, and expect their solution to work with all implementations of the Java platform.

Role of the JNI

When the Java platform is deployed on top of host environments, it may become desirable or necessary to allow Java applications to work closely with native code written in other languages. Programmers have begun to adopt the Java platform to build applications that were traditionally written in C and C++. Because of the existing investment in legacy code, however, Java applications will coexist with C and C++ code for many years to come. The JNI is a powerful feature that allows you to take advantage of the Java platform, but still utilize code written in other languages. As a part of the Java virtual machine implementation, the JNI is a *two-way* interface that allows Java applications to invoke native code and vice versa. Εικόνα 1.1 illustrates the role of the JNI.

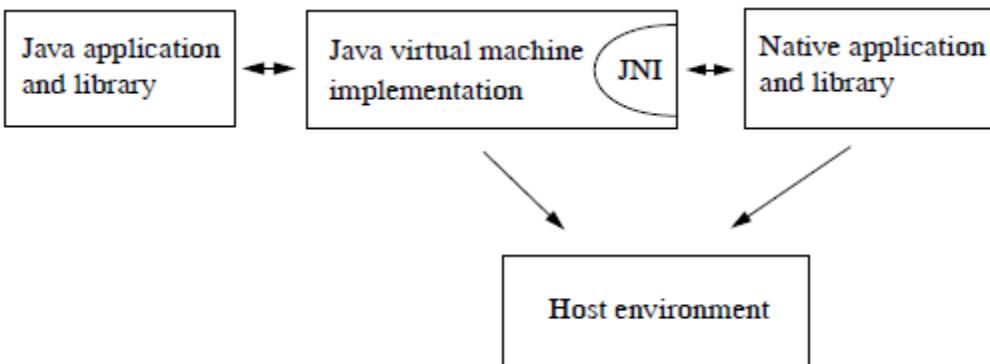


Figure 1.1 Role of the JNI

The JNI is designed to handle situations where you need to combine Java applications with native code. As a two-way interface, the JNI can support two types of native code: native *libraries* and native *applications*. • You can use the JNI to write *native methods* that allow Java applications to call functions implemented in native libraries. Java applications call native methods in the same way that they call methods implemented in the Java programming language. Behind the scenes, however, native methods are implemented in another language and reside in native libraries. • The JNI supports an *invocation interface* that allows you to embed a Java virtual machine

implementation into native applications. Native applications can link with a native library that implements the Java virtual machine, and then use the invocation interface to execute software components written in the Java programming language. For example, a web browser written in C can execute downloaded applets in an embedded Java virtual machine implementation.

9 Παράρτημα II

9.1 Windows library reference - Shell32 (provides access to system folder applications).

9.1.1 Shell32 library

The shell32 manages the Windows graphical shell. It implements a lot of the Windows Explorer. Some important shell32 concepts:

- The Shell Namespace. Like files and directories forms a filesystem hierarchy, the hierarchy displayed by shell32 is called the shell namespace. It contains all the files and directories as descendants of My Computer (and under wine the Unixfs '/' folder) but it contains many more folders (e.g. Control Panel, Network Neighborhood etc.)
- PIDs. Like files and directories are identified by paths that are sequences of names separated by '/' or '\', shell namespace items are identified by PIDs. PIDs are sequences of binary data structures. These structures should be opaque - that means that only the parent folder knows the binary data means. The user shouldn't try to interpret this data. However there are some applications that accesses this data (as far as I remember Word 97 is an example) so wine tries to make these structures compatible with Windows. A single structure is called an Item ID and a PID stands for Pointer to ID List. There are absolute and relative PIDs - exactly like with paths.
- Shell objects. In the filesystem we have a given set of operations that we can perform on a file. The shell namespace items are so diverse that there is no single set of operations. Instead we can query for COM objects associated with a given shell item. The most important interface is the IShellFolder. It allows to query for objects for it's subitems. Starting from the desktop folder obtained with SHGetDesktopFolder() we can obtain any folder in the shell namespace. Most folders implements also IShellFolder2. Some other important interfaces are IContextMenu, IDataObject (returns the path or PID of the item - used e.g. during Copy/Paste, when creating a link etc), IExtractIcon
- Shell views. Every IShellFolder can create a special IShellView object that manages a control that displays the contents of the folder. Shell32 allows to use any control to show the contents if one writes a custom IShellView. However usually the SHCreateShellFolderView

functions is used that create an IShellView with the listview that is well known to any Explorer folders.

- Shell browsers. A program that wants to display a shell view implements a shell browser. This interface is implemented by the Windows Explorer and by the Open/Save As common dialog.

9.1.2 User32 facilitates the simulation of the key strokes

9.2 Com4J - automatically generates the java classes that map directly to com interfaces

9.3 HWND - reference - provided by the com interface

9.4 win32GDI - to gain access to the display buffer