

Κατάτμηση Εικόνας με Μεθόδους Ομαδοποίησης Υπερ-εικονοστοιχείων

Κωνσταντίνα Λουκοπούλου

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ



Ιωάννινα, Φεβρουάριος 2010



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF IOANNINA



ΚΑΤΑΤΜΗΣΗ ΕΙΚΟΝΑΣ ΜΕ ΜΕΘΟΔΟΥΣ ΟΜΑΔΟΠΟΙΗΣΗΣ ΥΠΕΡ-ΕΙΚΟΝΟΣΤΟΙΧΕΙΩΝ

Η
ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

Υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύνθεσης
του Τμήματος Πληροφορικής
Εξεταστική Επιτροπή

από την

Κωνσταντίνα Λουκοπούλου

ως μέρος των Υποχρεώσεων

για τη λήψη

του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ
ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ - ΕΦΑΡΜΟΓΕΣ

Φεβρουάριος 2010



ΑΦΙΕΡΩΣΗ

*Στους αγαπημένους μου γονείς,
στην αδερφή μου Γιώτα,
στο Γιώργο...
και στο νομό μου, που από εκεί που βρίσκεται μου συμπαραστέκεται και με βοηθάει...*



ΕΥΧΑΡΙΣΤΙΕΣ

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα της εργασίας μου, κ. Χριστόφορο Νίκου για την καθοδήγηση και τις χρήσιμες παρατηρήσεις του.

Ένα μεγάλο ευχαριστώ οφείλω στο Δημήτρη Γερογιάννη και Γιώργο Σφήκα για την πολύτιμη συμβολή τους κατά την εκπόνηση της διατριβής.

Ευχαριστώ από τα βάθη της καρδιάς μου, τους γονείς μου, Θανάση και Βάσω, για την ηθική, ψυχολογική και οικονομική υποστήριξη που μου παρείχαν όλα αυτά τα χρόνια, την αδελφή μου Γιώτα, το θείο μου Μάκη και τη γιαγιά μου Γιώτα, για την απλόχερη αγάπη και το αδιάλειπτο ενδιαφέρον τους.

Ιδιαίτερος θα ήθελα να ευχαριστήσω το Γιώργο Φίλο για την ενθάρρυνση και την ηθική συμπαράσταση σε κάθε δυσκολία, αλλά και για την κατανόηση και την υπομονή του κατά τη συγγραφή της διατριβής αυτής.

Τέλος, θα ήθελα να ευχαριστήσω τη Μαρίνα Πλησίτη και το Θεοδόση Γκάμα για την ηθική υποστήριξη, αλλά και για τις αμέτρητες ευχάριστες στιγμές που μοιραστήκαμε στο ίδιο γραφείο καθώς επίσης και όλους τους φίλους μου που στάθηκαν δίπλα μου παρ' όλες τις δυσκολίες.



ΠΕΡΙΕΧΟΜΕΝΑ

	ΣΕΛ
ΑΦΙΕΡΩΣΗ	II
ΕΥΧΑΡΙΣΤΙΕΣ	III
ΠΕΡΙΕΧΟΜΕΝΑ	IV
ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ	VI
ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ	VIII
ΠΕΡΙΛΗΨΗ	XI
EXTENDED ABSTRACT IN ENGLISH	XII
ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ	1
1.1. Στόχοι	1
1.2. Δομή της Διατριβής	2
ΚΕΦΑΛΑΙΟ 2. ΥΠΟΛΟΓΙΣΜΟΣ ΥΠΕΡΕΙΚΟΝΟΣΤΟΙΧΕΙΩΝ	3
2.1. Τι Είναι τα Υπερ-εικονοστοιχεία;	3
2.2. Μέθοδοι για τον Υπολογισμό των Υπερ-εικονοστοιχείων	4
2.2.1. Πιθανοτικά Όρια Περιοχών	4
2.2.2. Αλγόριθμος N-Cut	9
2.2.3. Χαρακτηριστικά MRF	15
2.3. Ανάλυση σε Πρωτεύουσες Συνιστώσες	15
2.4. Ομαδοποίηση Δεδομένων με τον Αλγόριθμο K-μέσων	18
2.5. Πειραματικά Αποτελέσματα	21
2.5.1. Δείκτες Αξιολόγησης της Κατάτμησης	27
ΚΕΦΑΛΑΙΟ 3. ΚΑΤΑΤΜΗΣΗ ΕΙΚΟΝΑΣ ΜΕ ΤΗ ΘΕΩΡΙΑ ΡΥΘΜΟΥ - ΠΑΡΑΜΟΡΦΩΣΗΣ	37
3.1. Εισαγωγή	37
3.2. Θεωρία Ρυθμού – Παραμόρφωσης	37



3.3. Η Κανονική Κατανομή	40
3.4. Η Κατανομή Student's-t	41
3.5. Κατάτμηση Φυσικών Εικόνων χωρίς Επίβλεψη	43
3.6. Πειραματικά Αποτελέσματα	50
ΚΕΦΑΛΑΙΟ 4. ΕΠΙΛΟΓΟΣ	60
ΑΝΑΦΟΡΕΣ	62
ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ	64



ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας	Σελ
Πίνακας 2.1: Αποτελέσματα κατάτμησης εικόνων με χρήση <i>MRF</i> -χαρακτηριστικών και ύστερα ομαδοποίηση των μέσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 5$ περιοχές	29
Πίνακας 2.2: Αποτελέσματα κατάτμησης εικόνων με χρήση <i>MRF</i> -χαρακτηριστικών και ύστερα ομαδοποίηση των μέσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 10$ περιοχές	30
Πίνακας 2.3: Αποτελέσματα κατάτμησης εικόνων με χρήση <i>MRF</i> -χαρακτηριστικών και ύστερα ομαδοποίηση των μέσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 15$ περιοχές	30
Πίνακας 2.4: Αποτελέσματα κατάτμησης εικόνων με χρήση <i>MRF</i> -χαρακτηριστικών και ύστερα ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 5$ περιοχές	31
Πίνακας 2.5: Αποτελέσματα κατάτμησης εικόνων με χρήση <i>MRF</i> -χαρακτηριστικών και ύστερα ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 10$ περιοχές	31
Πίνακας 2.6: Αποτελέσματα κατάτμησης εικόνων με χρήση <i>MRF</i> -χαρακτηριστικών και ύστερα ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 15$ περιοχές	32
Πίνακας 2.7: Αποτελέσματα κατάτμησης εικόνων με χρήση <i>RGB</i> συντεταγμένων και ύστερα ομαδοποίηση των μέσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 5$ περιοχές	33
Πίνακας 2.8: Αποτελέσματα κατάτμησης εικόνων με χρήση <i>RGB</i> συντεταγμένων και ύστερα ομαδοποίηση των μέσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 10$ περιοχές	33
Πίνακας 2.9: Αποτελέσματα κατάτμησης εικόνων με χρήση <i>RGB</i> συντεταγμένων και ύστερα ομαδοποίηση των μέσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 15$ περιοχές	34
Πίνακας 2.10: Αποτελέσματα κατάτμησης εικόνων με χρήση <i>RGB</i> συντεταγμένων και ύστερα ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 5$ περιοχές	34
Πίνακας 2.11: Αποτελέσματα κατάτμησης εικόνων με χρήση <i>RGB</i> συντεταγμένων και ύστερα ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 10$ περιοχές	35
Πίνακας 2.12: Αποτελέσματα κατάτμησης εικόνων με χρήση <i>RGB</i> συντεταγμένων και ύστερα ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 15$ περιοχές	35
Πίνακας 3.1: Δείκτες αξιολόγησης με τη χρήση της Κανονικής κατανομής	52
Πίνακας 3.2: Δείκτες αξιολόγησης με τη χρήση της κατανομής Student's-t	53



Πίνακας 3.3: Δείκτες αξιολόγησης με τη χρήση της Κανονικής κατανομής	57
Πίνακας 3.4: Δείκτες αξιολόγησης με τη χρήση της κατανομής Student's-t	58



ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

Σχήμα	Σελ
Σχήμα 2.1: Κατατμήσεις τριών διαφορετικών αρχικών εικόνων από τρεις διαφορετικούς ανθρώπους	5
Σχήμα 2.2: Παραδείγματα εικονοστοιχειών με και χωρίς σύνορα	5
Σχήμα 2.3: Σχηματική αναπαράσταση του υπολογισμού των πιθανοτικών ορίων περιοχών	6
Σχήμα 2.4: Κύκλος κέντρου (x,y) , ακτίνας r και προσανατολισμού θ που δημιουργείται γύρω από κάθε εικονοστοιχείο	7
Σχήμα 2.5: Παράδειγμα ανίχνευσης υφής του φορέματος σε σχέση με το χέρι	7
Σχήμα 2.6: Καμπύλη Precision-Recall. Στόχος είναι η κατά το δυνατόν μεγιστοποίηση όχι μόνο του Precision αλλά και του Recall	8
Σχήμα 2.7: Σχηματική αναπαράσταση ενός γράφου	9
Σχήμα 2.8: Δημιουργία συνεκτικών τμημάτων ύστερα από τη διάσπαση των ακμών με μικρό βάρος	10
Σχήμα 2.9: Περίπτωση όπου η ελαχιστοποίηση του κριτηρίου $cut(A,B)$ επιτυγχάνεται με μια όχι τόσο «καλή» διάσπαση του γράφου, ενώ σε μια άλλη πιθανή διάσπαση το αποτέλεσμα είναι πιο αντιπροσωπευτικό χωρίς να ελαχιστοποιείται το κριτήριο	12
Σχήμα 2.10: Παράδειγμα εύρεσης του άξονα μέγιστης μεταβλητότητας με τη μέθοδο της <i>Ανάλυσης σε Πρωτεύουσες Συνιστώσες</i>	18
Σχήμα 2.11: Η εξάρτηση του K -μέσων από την αρχική διαμέριση	21
Σχήμα 2.12: α) Αρχική εικόνα, β) Υπερεικονοστοιχεία	22
Σχήμα 2.13: Όρια περιοχών που προκύπτουν για $K=5$ κέντρα με τη χρήση των RGB -συντεταγμένων και ύστερα από ομαδοποίηση των μέσων των υπερεικονοστοιχείων	22
Σχήμα 2.14: Όρια περιοχών που προκύπτουν για $K=10$ κέντρα με τη χρήση των RGB -συντεταγμένων και ύστερα από ομαδοποίηση των μέσων των υπερεικονοστοιχείων	22
Σχήμα 2.15: Όρια περιοχών που προκύπτουν για $K=15$ κέντρα με τη χρήση των RGB -συντεταγμένων και ύστερα από ομαδοποίηση των μέσων των υπερεικονοστοιχείων	23
Σχήμα 2.16: Όρια περιοχών που προκύπτουν για $K=5$ κέντρα με τη χρήση των χαρακτηριστικών MRF και ύστερα από ομαδοποίηση των μέσων των υπερεικονοστοιχείων	23
Σχήμα 2.17: Όρια περιοχών που προκύπτουν για $K=10$ κέντρα με τη χρήση των χαρακτηριστικών MRF και ύστερα από ομαδοποίηση των μέσων των υπερεικονοστοιχείων	24



Σχήμα 2.18: Όρια περιοχών που προκύπτουν για $K=15$ κέντρα με τη χρήση των χαρακτηριστικών MRF και ύστερα από ομαδοποίηση των μέσων των υπερεικονοστοιχείων	24
Σχήμα 2.19: Όρια περιοχών που προκύπτουν για $K=5$ κέντρα με τη χρήση των RGB -συντεταγμένων και ύστερα από ομαδοποίηση των medoid των υπερεικονοστοιχείων	25
Σχήμα 2.20: Όρια περιοχών που προκύπτουν για $K=10$ κέντρα με τη χρήση των RGB -συντεταγμένων και ύστερα από ομαδοποίηση των medoid των υπερεικονοστοιχείων	25
Σχήμα 2.21: Όρια περιοχών που προκύπτουν για $K=15$ κέντρα με τη χρήση των RGB -συντεταγμένων και ύστερα από ομαδοποίηση των medoid των υπερεικονοστοιχείων	25
Σχήμα 2.22: Όρια περιοχών που προκύπτουν για $K=5$ κέντρα με τη χρήση των χαρακτηριστικών MRF και ύστερα από ομαδοποίηση των medoid των υπερεικονοστοιχείων	26
Σχήμα 2.23: Όρια περιοχών που προκύπτουν για $K=10$ κέντρα με τη χρήση των χαρακτηριστικών MRF και ύστερα από ομαδοποίηση των medoid των υπερεικονοστοιχείων	26
Σχήμα 2.24: Όρια περιοχών που προκύπτουν για $K=15$ κέντρα με τη χρήση των χαρακτηριστικών MRF και ύστερα από ομαδοποίηση των medoid των υπερεικονοστοιχείων	27
Σχήμα 3.1: Γενική μορφή της γραφικής παράστασης Ρυθμού – Παραμόρφωσης	38
Σχήμα 3.2: Γραφική παράσταση πυκνότητας πιθανότητας Κανονικής κατανομής. Όσο μεγαλύτερη είναι η διασπορά, τόσο πιο πολύ απλώνονται τα άκρα της	40
Σχήμα 3.3: Γραφική παράσταση της συνάρτησης πυκνότητας πιθανότητας Κανονικής (κυανό) και Student's- t (κόκκινο)	41
Σχήμα 3.4: Γραφική παράσταση της συνάρτησης πυκνότητας πιθανότητας της Student's- t κατανομής για διάφορες τιμές της παραμέτρου ν . Παρατηρήστε ότι καθώς το ν μεγαλώνει, η καμπύλη προσεγγίζει αυτή της Κανονικής κατανομής	42
Σχήμα 3.5: Αλγόριθμος κατάτμησης με συμπίεση χωρίς απώλειες με την Κανονική κατανομή	49
Σχήμα 3.6: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's- t κατανομή	50
Σχήμα 3.7: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's- t κατανομή	50
Σχήμα 3.8: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's- t κατανομή.	51
Σχήμα 3.9: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's- t κατανομή	51
Σχήμα 3.10: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's- t κατανομή	52
Σχήμα 3.11: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's- t κατανομή	53
Σχήμα 3.12: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's- t κατανομή	54
Σχήμα 3.13: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's- t κατανομή	54



Σχήμα 3.14:	α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή	54
Σχήμα 3.15:	α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή	55
Σχήμα 3.16:	α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή	55
Σχήμα 3.17:	α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή	55
Σχήμα 3.18:	α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή	56
Σχήμα 3.19:	α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή	56
Σχήμα 3.20:	α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή	57



ΠΕΡΙΛΗΨΗ

Κωνσταντίνα Λουκοπούλου του Αθανασίου και της Βασιλικής. MSc, Τμήμα Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Φεβρουάριος, 2010. Κατάτμηση εικόνας με μεθόδους ομαδοποίησης υπερ-εικονοστοιχείων.

Επιβλέπων: Χριστόφορος Νίκου.

Τα υπερ-εικονοστοιχεία (super-pixels) είναι ομάδες από γειτονικά εικονοστοιχεία (pixels) με συγγενή χαρακτηριστικά ως προς το χρώμα και την υφή και παρέχουν μία υπερκατάτμηση της εικόνας. Στην εργασία αυτή παρουσιάζεται μία μέθοδος για την ομαδοποίηση υπερεικονοστοιχείων με σκοπό την κατάτμηση της εικόνας σε ένα μικρό αριθμό περιοχών. Αρχικά, τα δεδομένα κάθε υπερεικονοστοιχείου αναπαρίστανται με πολυδιάστατα διανύσματα για την περιγραφή τόσο της υφής όσο και του χρώματος και στη συνέχεια, με ανάλυση σε πρωτεύουσες συνιστώσες τα διανύσματα αυτά προβάλλονται σε χώρο μικρότερης διάστασης. Για κάθε υπερεικονοστοιχείο, τα νέα αυτά διανύσματα μοντελοποιούνται από μία συνάρτηση πυκνότητας πιθανότητας και η κατάτμηση της εικόνας επιτυγχάνεται με βάση τη θεωρία ρυθμού-παραμόρφωσης, συγχωνεύοντας επαναληπτικά εκείνα τα υπερεικονοστοιχεία για τα οποία ο ρυθμός μετάδοσης της πληροφορίας που περιέχει η ένωσή τους είναι μέγιστος για δεδομένη παραμόρφωση. Σε αυτό το πλαίσιο, εξετάστηκε η μοντελοποίηση κάθε υπερεικονοστοιχείου από μία κανονική κατανομή αλλά και από την εύρωστη κατανομή Student's-*t*. Η μέθοδος αξιολογήθηκε στη βάση φυσικών εικόνων του Berkeley με κριτήρια αξιολόγησης που συγκρίνουν τα αποτελέσματα των εξεταζόμενων τεχνικών με κατατιμήσεις εθελοντών.



EXTENDED ABSTRACT IN ENGLISH

Loukopoulou, Konstantina, MSc, Computer Science Department, University of Ioannina, Greece. February, 2010. Image segmentation by super-pixels merging
Supervisor: Christoforos Nikou

Super-pixels are groups of adjacent pixels with similar colour and texture attribute which provide an over-segmentation of an image. In this work, we present a method for clustering super-pixels, in order to segment an image in a small number of regions. Initially, each super-pixel's data are represented with multidimensional vectors for the description of texture and colour. For each super-pixel, these new vectors are modeled with a probability density function. Image segmentation is achieved using rate-distortion theory, by merging iteratively those super-pixels, for which the information rate after merging is maximum for a given distortion. Different modeling of superpixel multidimensional vector representation was studied. At first, a Gaussian distribution was employed for pixel modeling. A second approach using the robust Student's-*t* distribution was also considered. The segmentation methods were evaluated using the Berkeley image database using performance measures which compare the results of the techniques studied here with human segmentations.



ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

1.1 Στόχοι

1.2 Δομή της Διατριβής

1.1. Στόχοι

Στη συγκεκριμένη εργασία περιγράφεται μια μέθοδος με την οποία μπορούμε να ομαδοποιήσουμε τα εικονοστοιχεία μιας φυσικής εικόνας έτσι ώστε οι ομάδες που δημιουργούνται να έχουν κάποια φυσική σημασία. Αρχικά χρησιμοποιείται η έννοια των υπερ-εικονοστοιχείων που είναι ομάδες γειτονικών εικονοστοιχείων με παρόμοια χαρακτηριστικά ως προς το χρώμα και την υφή. Επειδή τα υπερ-εικονοστοιχεία παρέχουν μια υπερκατάτμηση της εικόνας, με τη χρήση μεθόδων ρυθμού-παραμόρφωσης, συνενώνουμε γειτονικά υπερ-εικονοστοιχεία, με σκοπό την αναγνώριση των αντικειμένων της εικόνας. Η αναπαράσταση των εικονοστοιχείων γίνεται με πολυδιάστατα διανύσματα. Επειδή η διάσταση των διανυσμάτων αυτών είναι μεγάλη, χρησιμοποιούμε τη μέθοδο της ανάλυσης σε πρωτεύουσες συνιστώσες, η οποία προβάλλει τα δεδομένα μας στον άξονα μέγιστης συμμεταβλητότητας, με αποτέλεσμα την προβολή τους σε χώρο μικρότερης διάστασης. Η μοντελοποίηση των δεδομένων γίνεται με τη χρήση κάποιων κατανομών, της κανονικής κατανομής, καθώς επίσης και της κατανομής Student's-*t*. Η κατάτμηση τελικά, επιτυγχάνεται με βάση τη θεωρία ρυθμού - παραμόρφωσης, συγχωνεύοντας επαναληπτικά τα υπερεικονοστοιχεία για τα οποία ο ρυθμός μετάδοσης της πληροφορίας που περιέχει η ένωσή τους είναι μέγιστος, δεδομένης κάποιας παραμόρφωσης. Στόχος της εργασίας αυτής, είναι η σύγκριση των κατατμήσεων που προκύπτουν με τη χρήση της κατανομής Student's-*t* με την κανονική κατανομή. Θεωρητικά, με τη χρήση της Student's-*t*, αναμένουμε τα αποτελέσματα της κατάτμησης της εικόνας να είναι καλύτερα σε σχέση με τη χρήση της κανονικής κατανομής, και αυτό αιτιολογείται από τη μορφή της κατανομής αυτής, από τις πιθανότητες



δηλαδή που δίνει στα εικονοστοιχεία να ανήκουν ή όχι στην ίδια ομάδα. Τα κριτήρια αξιολόγησης που χρησιμοποιήθηκαν για τη σύγκριση των κατατμήσεων που προέκυψαν σε σχέση με ανθρώπινες κατατμήσεις, είναι το Rand Index, το Boundary Detection Error, το Global Consistency Error και το Variation of Information. Για την αξιολόγηση της μεθόδου που περιγράψαμε, χρησιμοποιήθηκε η βάση φυσικών εικόνων του Berkeley.

1.2. Δομή της Διατριβής

Η διατριβή περιέχει τα εξής κεφάλαια: Στο Κεφάλαιο 1 παρουσιάζονται οι στόχοι της εργασίας. Στο Κεφάλαιο 2 συμπεριλαμβάνεται το θεωρητικό υπόβαθρο και κάποιες βασικές μέθοδοι για τον υπολογισμό των υπερ-εικονοστοιχείων. Στο Κεφάλαιο 3 προτείνεται μια μέθοδος κατάτμησης φυσικών εικόνων με τη χρήση της θεωρίας ρυθμού-παραμόρφωσης και τη μοντελοποίηση από την κανονική κατανομή και την εύρωστη κατανομή Student's-t. Στο Κεφάλαιο 4 συμπεριλαμβάνεται ο επίλογος της διατριβής.



ΚΕΦΑΛΑΙΟ 2. ΥΠΟΛΟΓΙΣΜΟΣ ΥΠΕΡΕΙΚΟΝΟ- ΣΤΟΙΧΕΙΩΝ

-
- 2.1 Τι Είναι τα Υπερ-εικονοστοιχεία;
 - 2.2 Μέθοδοι για τον Υπολογισμό των Υπερ-εικονοστοιχείων
 - 2.3 Ομαδοποίηση Δεδομένων με τον Αλγόριθμο K-μέσων
 - 2.4 Ανάλυση σε Πρωτεύουσες Συνιστώσες
 - 2.5 Πειραματικά Αποτελέσματα
-

2.1. Τι Είναι τα Υπερ-εικονοστοιχεία;

Η κατάτμηση εικόνας είναι ένα από τα σημαντικότερα και δυσκολότερα προβλήματα στον τομέα της όρασης υπολογιστών. Γενικά, είναι η διαδικασία απομόνωσης των αντικειμένων στην εικόνα από το φόντο, δηλαδή χωρισμού της εικόνας σε περιοχές που είναι συνδεδεμένες και ομοιογενείς με βάση κάποιο κριτήριο, όπως το επίπεδο του γκρι, η υφή, το χρώμα, η κίνηση και το βάθος. Εφόσον ο όρος «ομοιογενής» είναι μάλλον ασαφής και επιτρέπει πολλές διαφορετικές ερμηνείες της έννοιάς του, ο στόχος της κατάτμησης εικόνας εμφανίζεται να είναι εξαρτώμενος τόσο από το κριτήριο που εφαρμόζεται, όσο και από την εφαρμογή. Ακόμη κι ένας άνθρωπος μπορεί να αντιμετωπίσει μερικές σχετικές δυσκολίες τμηματοποιώντας μια εικόνα, και μπορούν να προκύψουν ποικίλες διαφορετικές κατατμήσεις από διαφορετικούς ανθρώπινους παρατηρητές όπως φαίνεται στο Σχήμα 2.1.

Τα υπερ-εικονοστοιχεία (super-pixels) αποτελούν μία βασική έννοια που έχει απασχολήσει έντονα την επιστημονική κοινότητα. Ως υπερ-εικονοστοιχείο ορίζεται ένα σύνολο από κάποια εικονοστοιχεία μιας εικόνας που είναι γειτονικά μεταξύ τους



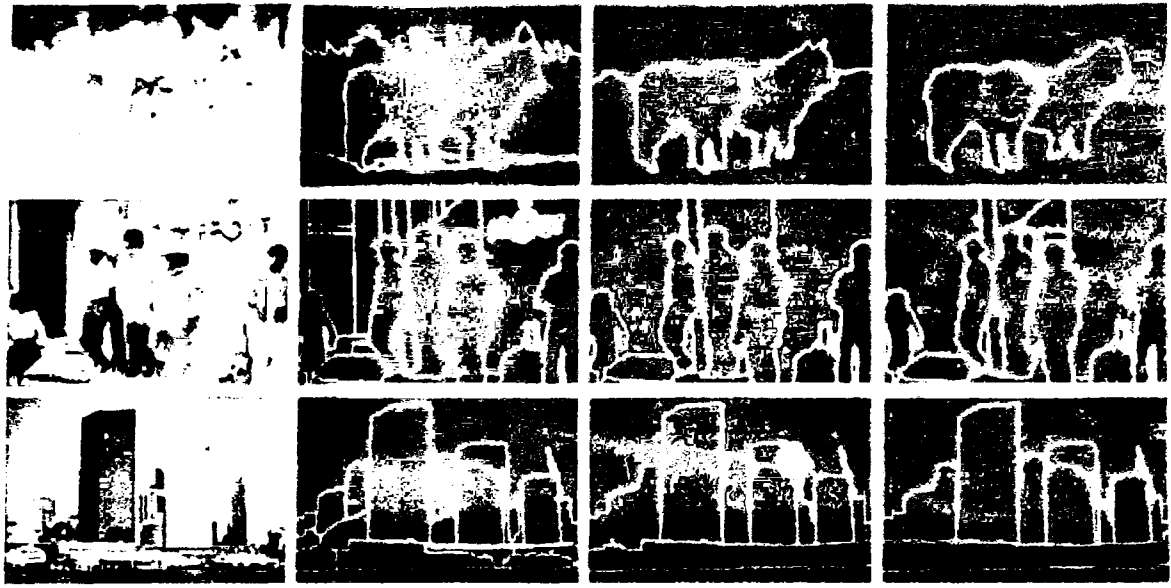
και μοιάζουν σε κάποια χαρακτηριστικά τους - όπως για παράδειγμα στη φωτεινότητα, το χρώμα και την υφή τους. Ένα εικονοστοιχείο δεν μπορεί να ανήκει σε πολλά υπερ-εικονοστοιχεία ταυτόχρονα, πράγμα το οποίο σημαίνει ότι τα υπερ-εικονοστοιχεία δεν είναι επικαλυπτόμενα μεταξύ τους και η ένωση όλων των τμημάτων αυτών, μας δίνει ολόκληρη την εικόνα. Έτσι, τα εικονοστοιχεία μιας εικόνας μπορούν να ομαδοποιηθούν, έτσι ώστε τα pixels μιας ομάδας (ενός super-pixel), να είναι όμοια μεταξύ τους σε κάποια ή σε όλα τα χαρακτηριστικά που αναφέραμε παραπάνω. Τα υπερ-εικονοστοιχεία τελικά, χωρίζουν την εικόνα μας σε περιοχές στις οποίες τα όρια των «αντικειμένων» συμπίπτουν με τα υπερ-εικονοστοιχεία, αλλά ένα αντικείμενο αποτελείται από πολλά υπερεικονοστοιχεία.

2.2. Μέθοδοι για τον Υπολογισμό των Υπερ-εικονοστοιχείων

2.2.1. Πιθανοτικά Όρια Περιοχών

Σύμφωνα με το [4], η μέθοδος των πιθανοτικών ορίων περιοχών (Probabilistic Boundary) έχει στόχο να υπολογίζει όρια σε φυσικές εικόνες, χρησιμοποιώντας κάποιες τοπικές μετρήσεις. Τα χαρακτηριστικά που χρησιμοποιούνται είναι η φωτεινότητα, το χρώμα και την υφή. Για να συνδυάσουμε την πληροφορία που προκύπτει από τα παραπάνω χαρακτηριστικά, εκπαιδεύουμε έναν ταξινομητή χρησιμοποιώντας σαν γενικευμένη αλήθεια (ground truth) τμηματοποιήσεις μιας εικόνας από διάφορους ανθρώπους. Το αποτέλεσμα που θα πάρουμε σαν έξοδο από αυτό τον ταξινομητή, μας δίνει την πιθανότητα να έχουμε όριο στο συγκεκριμένο σημείο και με τη συγκεκριμένη κατεύθυνση του ορίου. Χρησιμοποιούνται καμπύλες Precision-Recall για να δείξουμε πόσο καλά είναι τα τελικά αποτελέσματα του εντοπισμού του ορίου σε εικονοστοιχεία σε σχέση με τα ανθρώπινα αποτελέσματα.



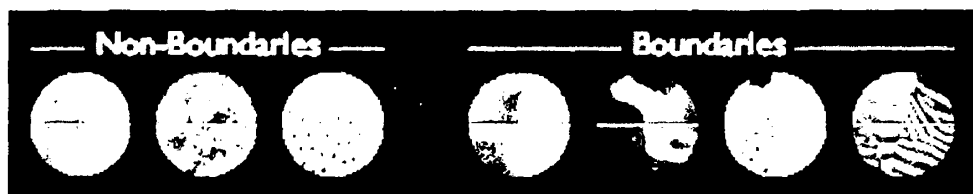


Σχήμα 2.1: Κατατμήσεις τριών διαφορετικών αρχικών εικόνων από τρεις διαφορετικούς ανθρώπους

Στο Σχήμα 2.1, βλέπουμε τις κατατμήσεις που έχουν γίνει από τρεις διαφορετικούς ανθρώπους για κάποιες εικόνες. Οι κατατμήσεις αυτές, είναι παρόμοιες, αλλά δεν είναι ίδιες και έχουν γίνει με βάση κάποια χαρακτηριστικά όπως, η φωτεινότητα, το χρώμα και η υφή των pixels όπως αναφέρθηκε και παραπάνω.

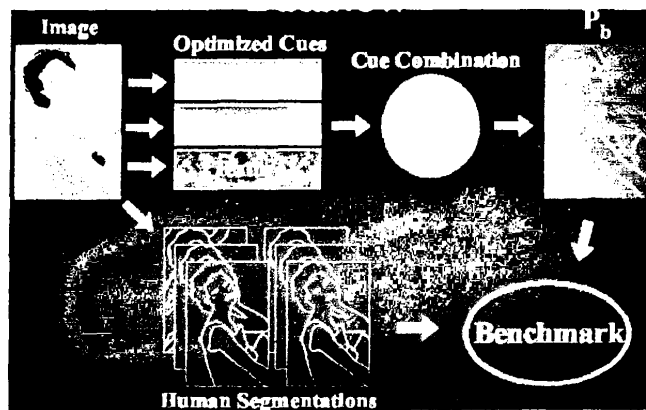
Στόχος του [4] είναι να βρεθεί ο βέλτιστος τρόπος χρήσης της τοπικής πληροφορίας που παίρνουμε από κάποια εικονοστοιχεία και τη γειτονιά τους.

Στο Σχήμα 2.2, φαίνονται κάποια παραδείγματα εικονοστοιχείων στα στα οποία υπάρχει ή δεν υπάρχει όριο, σύμφωνα με το χρώμα, τις *Lab* συντεταγμένες καθώς επίσης και τη συχνότητα, την κατεύθυνση και την αντίθεση που υπάρχει σε ένα εικονοστοιχείο. Η συχνότητα, η κατεύθυνση και η αντίθεση που υπάρχει σε ένα εικονοστοιχείο καθορίζουν τα όρια τα οποία τίθενται βάσει της υφής τους.



Σχήμα 2.2: Παραδείγματα εικονοστοιχειων με και χωρίς σύνορα

Από την εικόνα μας λοιπόν, παίρνουμε όλα τα βελτιστοποιημένα χαρακτηριστικά φωτεινότητας, χρώματος και υφής και ύστερα από το συνδυασμό τους, παράγουμε ένα μοντέλο. Στη συνέχεια, από το μοντέλο αυτό, δημιουργούμε ένα “χάρτη συνόρων”, που μας δείχνει τα όρια που παρήχθησαν για την αρχική εικόνα. Συγκρίνουμε τα αποτελέσματα του “χάρτη συνόρων” με αυτά των ανθρώπινων καταταμήσεων και ελέγχουμε κατά πόσο ταιριάζουν μεταξύ τους – και επομένως κατά πόσο σωστός ήταν ο “χάρτης συνόρων” που δημιουργήθηκε. Η διαδικασία που περιγράφηκε, φαίνεται στο Σχήμα 2.3.



Σχήμα 2.3: Σχηματική αναπαράσταση του υπολογισμού των πιθανοτικών ορίων περιοχών

Αναλυτικότερα, για να ανιχνεύσουμε τοπικές αλλαγές στο χρώμα, την υφή και τη φωτεινότητα σε κάποιο εικονοστοιχείο της εικόνας -έστω (x,y) -, σχεδιάζουμε έναν κύκλο ακτίνας r και τον χωρίζουμε κατά μήκος της διαμέτρου και με προσανατολισμό θ (δοκιμάζουμε πολλές διαφορετικές τιμές του r και του θ). Ακολούθως, συγκρίνουμε τα περιεχόμενα των δύο μισών του κύκλου. Μεγάλη αλλαγή στο χρώμα και τη φωτεινότητα δείχνει ότι κατά μήκος αυτής της διαμέτρου έχουμε σύνορο. Για την ανίχνευση της υφής, εφαρμόζουμε κάποια φίλτρα σε κάθε ημικύκλιο. Για τη σύγκριση των περιεχομένων των δύο μισών του κύκλου, χρησιμοποιούμε την ποσότητα:

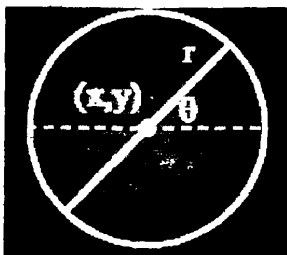
$$x^2(g, h) = \frac{1}{2} \sum_{i=1}^n \frac{(g_i - h_i)^2}{g_i + h_i} \quad (2.1)$$

όπου g και h οι κατανομές για καθένα από τα δύο ημικύκλια που προέκυψαν.

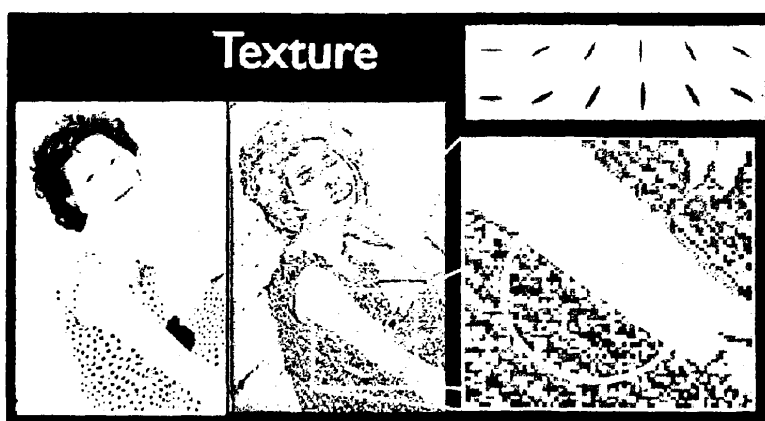
Αν η ποσότητα αυτή έχει μεγάλη τιμή, τότε θα υπάρχει σύνορο, αλλιώς όχι.

Στη συνέχεια, για να ελέγξουμε ποιες από τις ακμές που προέκυψαν από την παραπάνω διαδικασία είναι πραγματικές ακμές, συγκρίνουμε τα σύνορα που προέκυψαν με τα οπτικά αποτελέσματα που προκύπτουν από κατατμήσεις εθελοντών.

Στα σχήματα 2.4 και 2.5 μπορούμε να δούμε τη διαδικασία που μόλις περιγράφηκε.



Σχήμα 2.4: Κύκλος κέντρου (x,y) , ακτίνας r και προσανατολισμού θ που δημιουργείται γύρω από κάθε εικονοστοιχείο



Σχήμα 2.5: Παράδειγμα ανίχνευσης υφής του φορέματος σε σχέση με το χέρι

Το βήμα που ακολουθεί όπως είδαμε στο Σχήμα 2.3, είναι αυτό του συνδυασμού των βελτιστοποιημένων χαρακτηριστικών (του κάθε χαρακτηριστικού ξεχωριστά). Στη συνέχεια, συνδυάζουμε τα χαρακτηριστικά μεταξύ τους ώστε να παραχθεί το βέλτιστο δυνατό αποτέλεσμα.

Επειτα, με τη χρήση κάποιου ταξινομητή αποφασίζουμε αν τα σύνορα καταλήξαμε είναι σωστά ή όχι, και κατασκευάζουμε μία Precision-Recall καμπύλη όπου Precision (P) μας δείχνει πόσο θόρυβο μπορούμε να ανεχθούμε και ισχύει:

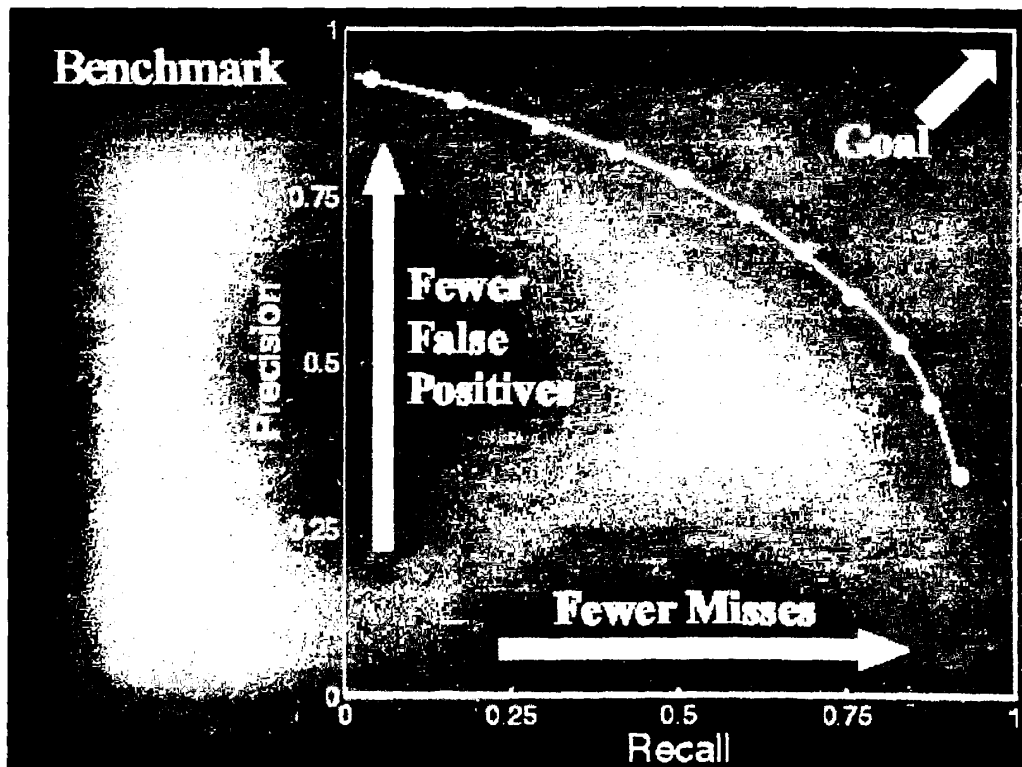
$$P = \frac{TP}{TP + FP}$$

όπου, TP : αληθές θετικό (true positive), FP : ψευδές θετικό (false positive), και Recall (R) μας δείχνει από τα συνολικά σύνορα που καταλήξαμε, ππραγματικά είναι και ισχύει:

$$R = \frac{TP}{TP + FN}$$

όπου FN : ψευδές αρνητικό (false negative).

Αυτή η μέθοδος μοντελοποιεί την εκ των υστέρων πιθανότητα να έχουμε σύνορα κάθε εικονοστοιχείο και προσανατολισμό, πράγμα το οποίο είναι χρήσιμο αλγορίθμους υψηλότερου επιπέδου.

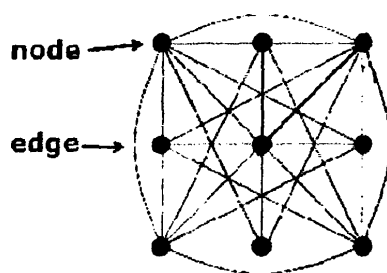


Σχήμα 2.6: Καμπύλη Precision-Recall. Στόχος είναι η κατά το δυναμειστοποίηση όχι μόνο του Precision αλλά και του Recall



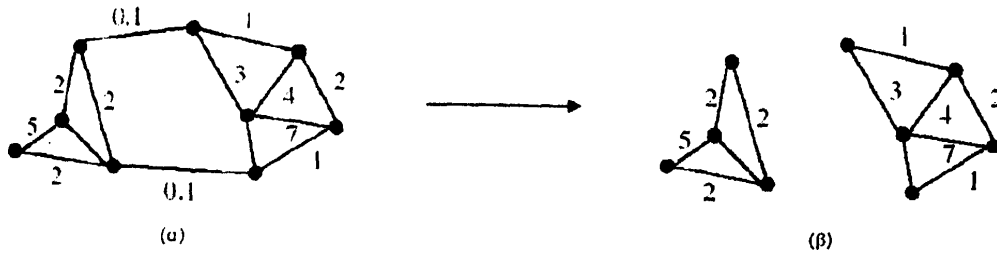
2.2.2. Αλγόριθμος *N-Cut*

Μία πολύ σημαντική τεχνική για τη λύση του προβλήματος της τμηματοποίησης μιας εικόνας αποτελεί η αντιμετώπισή του ως πρόβλημα διάσπασης ή σύνδεσης γράφου [9]. Γράφος G , είναι ένα σύνολο το οποίο αποτελείται από κορυφές V και ακμές E , δηλαδή $G = \{V, E\}$, Σχήμα 2.7. Κάθε ακμή ενώνει ένα ζεύγος κορυφών και ισχύει: $E \subset V \times V$. Δύο κορυφές ενός γράφου ενώνονται όταν υπάρχει μια σειρά ακμών με αφετηρία τη μία κορυφή του γράφου και τέλος την άλλη. Σε κάθε ακμή ενός ζεύγους κορυφών v_i, v_j αντιστοιχεί μια αριθμητική τιμή $W(i, j)$ που ονομάζεται βάρος του γράφου και υποδηλώνει την ομοιότητα των κορυφών i, j .



Σχήμα 2.7: Σχηματική αναπαράσταση ενός γράφου

Στο πρόβλημα τμηματοποίησης εικόνας με γράφο, τα εικονοστοιχεία της εικόνας αποτελούν τις κορυφές του γράφου και κάθε ακμή (μη κατευθυνόμενη), έχει βάρος το οποίο όσο μεγαλύτερο είναι, τόσο μεγαλύτερη ομοιότητα έχουν τα δύο εικονοστοιχεία που συνδέει. Κάθε κόμβος του γράφου μπορεί να αποτελείται από πολλά εικονοστοιχεία τα οποία αποτελούν μια περιοχή της εικόνας. Τα τμήματα αυτά ονομάζονται συνεκτικά (connected components). Σε μια τέτοια αντιμετώπιση του προβλήματος της τμηματοποίησης μιας εικόνας ο στόχος είναι η διάσπαση του γράφου σε τόσα συνεκτικά τμήματα όσα είναι και τα φυσικά αντικείμενα της εικόνας. Εάν ο γράφος διασπαστεί στις ακμές όπου η τιμή του βάρους είναι μικρή, τότε τα συνεκτικά τμήματα που δημιουργούνται αποτελούνται από κόμβους με μεγάλη ομοιότητα.



Σχήμα 2.8: Δημιουργία συνεκτικών τμημάτων ύστερα από τη διάσπαση των ακμών με μικρό βάρος

Στο Σχήμα 2.8(α) βλέπουμε ένα γράφο με βάρη που χαρακτηρίζουν την ομοιότητα ανάμεσα στους κόμβους. Εάν διασπάσουμε τις ακμές με μικρό βάρος, παρατηρούμε ότι δημιουργούνται τα συνεκτικά τμήματα που φαίνονται στο Σχήμα 2.8(β).

Ένας γράφος αντιπροσωπεύεται από έναν τετραγωνικό πίνακα, τον πίνακα βαρών. Σ' αυτόν τον πίνακα, οι μεγάλες τιμές βαρών είναι φωτεινές, ενώ οι μικρές είναι σκοτεινές.

Η μεθοδολογία που χρησιμοποιείται για την κατασκευή του πίνακα βαρών, είναι η εξής: Αρχικά, παίρνουμε κάθε διδιάστατο σημείο της συλλογής και το αντιστοιχίζουμε με μια κορυφή [12]. Στη συνέχεια, συνδέουμε κάθε στοιχείο με όλα τα άλλα στοιχεία με μια ακμή. Το βάρος της ακμής εξαρτάται από την ομοιότητα των στοιχείων. Έπειτα, κόβουμε τις ακμές του γράφου οι οποίες έχουν μικρά βάρη για να φτιάξουμε ένα καλό σύνολο συγκεντρωτικών συστατικών. Κάθε συστατικό που δημιουργείται είναι και μια ομάδα.

Ο πίνακας βαρών δίνεται από τον παρακάτω τύπο:

$$W = e^{-\frac{\|F(i)-F(j)\|_2^2}{\sigma_f}} * \begin{cases} e^{-\frac{\|X(i)-X(j)\|_2^2}{\sigma_x}}, & \text{αν } \|X(i)-X(j)\|_2 < r \\ 0, & \text{αλλιώς} \end{cases} \quad (2.4)$$

όπου $X(i)$ είναι ένα διάνυσμα με την ακριβή θέση του κόμβου i στην αρχική μας εικόνα, ή οι ακριβείς συντεταγμένες ενός σημείου και $F(i) = I(i)$, δηλαδή είναι η τιμή της φωτεινότητας της αρχικής εικόνας στον κόμβο I , σ_1 και σ_x ελεύθερες παράμετροι και r ένα κατώφλι για την απόσταση με βάση το οποίο καθορίζεται το βάρος μεταξύ των κόμβων. Η σχέση 2.4 πολλαπλασιάζεται με έναν όρο ο οποίος μας δίνει την πιθανότητα να υπάρχει σύνορο ή όχι (πιθανοτικά όρια περιοχών).

Ο αλγόριθμος που χρησιμοποιείται κατά βάση για την κατάτμηση της εικόνας είναι ο *Ncut*. Η μέθοδος αυτή, μετρά την ανομοιότητα μεταξύ δύο διαφορετικών ομάδων, καθώς επίσης και την ομοιότητα των κόμβων (εικονοστοιχείων) που βρίσκονται στην ίδια ομάδα [9].

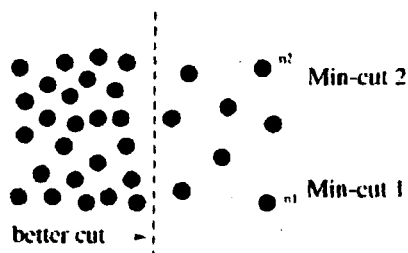
Ένας γράφος $G = \{V, E\}$ μπορεί να χωριστεί σε δύο μη συνεκτικά σύνολα A, B , $A \cup B = V$, $A \cap B = \emptyset$, αφαιρώντας απλά τις ακμές που συνδέουν τα δύο σύνολα. Ο βαθμός ανομοιότητας μεταξύ των δύο συνόλων μπορεί να υπολογιστεί ως το συνολικό βάρος των ακμών που αφαιρέθηκαν. Στη θεωρία των γράφων, αυτό ονομάζεται *cut*:

$$cut(A, B) = \sum_{u \in A, v \in B} w(u, v) \quad (2.5)$$

Δηλαδή το $cut(A, B)$ είναι το άθροισμα των βαρών όλων των ακμών του V , που έχουν έναν κόμβο στο A και έναν κόμβο στο B . Η βέλτιστη κατάτμηση ενός γράφου σε δύο υπογράφους είναι αυτή που ελαχιστοποιεί την τιμή της ποσότητας *cut*. Υπάρχουν όμως περιπτώσεις που η ελαχιστοποίηση του κριτηρίου αυτού δε συντελεί σε μια «καλή» τμηματοποίηση γιατί είναι δυνατό να απομονωθούν στο ελάχιστο κάποιοι κόμβοι του γράφου. Πιο αναλυτικά, η ελαχιστοποίηση του κριτηρίου επιτυγχάνεται με τη διάσπαση μικρών συνόλων από απομονωμένους κόμβους του γράφου. Εξ' ορισμού η τιμή του $cut(A, B)$ μεταβάλλεται ανάλογα με τον αριθμό των ακμών ανάμεσα στα δύο μέρη όπου πραγματοποιήθηκε η διάσπαση. Στο Σχήμα 2.9 παρουσιάζεται μια τέτοια περίπτωση. Στο παράδειγμα αυτό εάν υποθέσουμε ότι το βάρος των ακμών είναι αντιστρόφως ανάλογο με την απόσταση των δύο κόμβων, τότε με τη διάσπαση του n_1 ή n_2 το κριτήριο έχει πολύ μικρή τιμή, ενώ αν η διάσπαση



πραγματοποιούνταν σε κάποιο άλλο πιθανό σημείο όπως φαίνεται στο σχήμα, το αποτέλεσμα είναι αντιπροσωπευτικό χωρίς να ελαχιστοποιείται το κριτήριο.



Σχήμα 2.9: Περίπτωση όπου η ελαχιστοποίηση του κριτηρίου $cut(A,B)$ επιτυγχάνεται με μια όχι τόσο «καλή» διάσπαση του γράφου, ενώ σε μια άλλη πιθανή διάσπαση το αποτέλεσμα είναι πιο αντιπροσωπευτικό χωρίς να ελαχιστοποιείται το κριτήριο

Στην περίπτωση του αλγορίθμου $Ncut$ αντί να κοιτάμε την τιμή του συνολικού βάρους που συνδέει τις ακμές των δύο ομάδων, το μέτρο που χρησιμοποιούμε υπολογίζει το κόστος ως το άθροισμα των παρακάτω κλασμάτων, το οποίο κανονικοποιεί την τιμή του $cut(A,B)$ και δίνεται από τον παρακάτω τύπο:

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)} \quad (2.6)$$

όπου $assoc(A,V) = \sum_{u \in A, t \in V} w(u,t)$ και είναι οι συνολικές συνδέσεις από τους κόμβους

του A προς όλους τους κόμβους του γράφου, αντίστοιχα και για το $assoc(B,V)$. Η τιμή της παραπάνω συνάρτησης κόστους ($Ncut$) είναι μικρή αν η περικοπή χωρίζει δύο συστατικά που έχουν λίγες ακμές χαμηλού βάρους μεταξύ τους και πολλές εσωτερικές ακμές μεγάλου βάρους.

Στο ίδιο πνεύμα, ορίζεται ένα άλλο μέτρο για τη συνολική ομαλοποιημένη ένωση στις ομάδες:

$$Nassoc(A,B) = \frac{assoc(A,A)}{assoc(A,V)} + \frac{assoc(B,B)}{assoc(B,V)} \quad (2.7)$$

όπου $assoc(A, A)$ και $assoc(B, B)$ είναι τα συνολικά βάρη των ακμών που συνδέουν κόμβους που βρίσκονται μέσα στην ομάδα A και B αντίστοιχα. Αυτό είναι ένα μέτρο που απεικονίζει το πόσο στενά κατά μέσο όρο οι κόμβοι μέσα στην ίδια ομάδα συνδέονται ο ένας με τον άλλο. Από τα παραπάνω, έχουμε ότι:

$$\begin{aligned} Ncut(A, B) &= \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)} = \frac{assoc(A, V) - assoc(A, A)}{assoc(A, V)} + \frac{assoc(B, V) - assoc(B, B)}{assoc(B, V)} \\ &= 2 - \left(\frac{assoc(A, A)}{assoc(A, V)} + \frac{assoc(B, B)}{assoc(B, V)} \right) = 2 - Nassoc(A, B) \end{aligned}$$

Ως εκ τούτου, τα δύο κριτήρια χωρισμάτων που επιδιώκουμε στον αλγόριθμο ομαδοποίησης μας, ελαχιστοποιεί την αποσύνδεση μεταξύ των ομάδων και μεγιστοποιεί την ένωση μέσα στις ομάδες.

Ο αλγόριθμος $Ncut$ λειτουργεί ως εξής:

Θεωρεί τη συνολική σύνδεση από τον κόμβο i σε όλους τους άλλους κόμβους,

$$D(i, i) = \sum_j w(i, j) \quad \text{ή} \quad d(i) = \sum_j w(i, j) \quad (2.9)$$

Η μέθοδος αυτή, χρησιμοποιεί τα k μικρότερα ιδιοδιανύσματα των k μικρότερων ιδιοτιμών από το πρόβλημα ιδιοτιμών:

$$(D - W)y = \lambda Dy \quad (2.10)$$

Η διάσταση του πίνακα W είναι $(N \times M) \times (N \times M)$, αν το μέγεθος της αρχικής εικόνας είναι $N \times M$, ενώ το μέγεθος του D είναι $(N \times M) \times (N \times M)$. Η εξίσωση αυτή, μπορεί να γραφεί και ως:

$$D^{-1/2}(D - W)D^{-1/2}z = \lambda z,$$

όπου $z = D^{1/2}x$.



Μπορούμε εύκολα να δούμε, ότι το $z_0 = D^{1/2}1$ είναι ένα ιδιοδιάνυσμα της $D^{-1/2}(D-W)D^{-1/2}z = \lambda z$, αλλά με ιδιοτιμή ίση με μηδέν. Ο $(D-W)$ πίνακας, λέγεται αλλιώς «Λαπλασιανός» πίνακας. Το z_0 είναι το μικρότερο ιδιοδιάνυσμα, το z_1 το αμέσως μεγαλύτερο, κλπ. Επομένως, $y_0 = 1$ είναι το μικρότερο ιδιοδιάνυσμα με ιδιοτιμή 0 και $0 = z_1^T z_0 = y_1^T D1$, όπου y_1 είναι το δεύτερο μικρότερο ιδιοδιάνυσμα της $D^{-1/2}(D-W)D^{-1/2}z = \lambda z$.

Συνολικά, ο αλγόριθμος *Ncut* είναι ο παρακάτω:

- Για την αρχική εικόνα, φτιάχνουμε ένα γράφο βαρών $G = (V, E)$.
- Θεωρούμε κάθε εικονοστοιχείο ως έναν κόμβο και ενώνουμε κάθε κόμβο με όλους τους υπόλοιπους.
- Η ακμή μεταξύ δύο κόμβων θα έχει βάρος ανάλογα με την ομοιότητα αυτών των κόμβων.
- Όσο μεγαλύτερη είναι η ομοιότητα, τόσο μεγαλύτερο θα είναι και το βάρος της σύνδεσης. Όμοιοι κοντινοί κόμβοι, έχουν μεγάλο βάρος ακμής.

Με τον παραπάνω τρόπο, σχηματίζεται ο τετραγωνικός πίνακας βαρών W .

Λύνοντας το πρόβλημα ιδιοτιμών

$$(D-W)y = \lambda Dy \text{ ή το } D^{-1/2}(D-W)D^{-1/2}z = \lambda z,$$

όπου $z = D^{1/2}x$, -δηλαδή ελαχιστοποιώντας την ποσότητα *Ncut* υπό τη συνθήκη:

$$y_1^T D1 = 0$$

βρίσκουμε τις k μικρότερες ιδιοτιμές, και στη συνέχεια τα k μικρότερα ιδιοδιανύσματα που αντιστοιχούν σε αυτές τις ιδιοτιμές.

Εφαρμόζουμε κάποιον αλγόριθμο ομαδοποίησης (για παράδειγμα τον αλγόριθμο K -μέσων) στα K ιδιοδιανύσματα που βρήκαμε έτσι ώστε να κόψουμε το γράφο, και έτσι παίρνουμε K διαφορετικές ομάδες.



2.2.3. Χαρακτηριστικά *MRF*

Για την αναγνώριση και την ομαδοποίηση των εικονοστοιχείων μιας εικόνας με βάση την υφή, συνήθως χρησιμοποιούνται οι αποκρίσεις σε κάποια φίλτρα. Παρόλα αυτά, έχει αποδειχθεί [11] ότι η τράπεζα φίλτρων δεν είναι απαραίτητη, διότι μπορούν να χρησιμοποιηθούν τα *MRF*-χαρακτηριστικά της εικόνας, τα οποία την περιγράφουν εξίσου καλά με την εφαρμογή κάποιων φίλτρων.

Για κάθε εικονοστοιχείο παίρνουμε την τοπική γειτονιά του αντί για τις αποκρίσεις σε κάποια φίλτρα. Έχει αποδειχθεί ότι μικρές γειτονιές μεγέθους για παράδειγμα 7×7 μπορούν να μας δώσουν καλή απόδοση ταξινόμησης. Εάν το μέγεθος της γειτονιάς λοιπόν είναι 7×7 και εφόσον η εικόνα είναι έγχρωμη, δημιουργούμε ένα $7 \times 7 \times 3 = 147 \times 1$ διάνυσμα για κάθε εικονοστοιχείο της εικόνας. Έτσι, ο πίνακας συμμεταβλητότητας θα έχει μέγεθος 147×147 . Εάν εφαρμόσουμε τη μέθοδο της ανάλυσης σε πρωτεύουσες συνιστώσες και κρατήσουμε τα δ για παράδειγμα μεγαλύτερα διανύσματα, καταφέρνουμε να περιγράψουμε κάθε εικονοστοιχείο με ένα οκταδιάστατο διάνυσμα, το οποίο περιγράφει «καλά» την αρχική μας εικόνα.

Επομένως, η μέθοδος χρήσης των χαρακτηριστικών *MRF* είναι μια μέθοδος που μπορούμε να χρησιμοποιήσουμε για την ομαδοποίηση των εικονοστοιχείων με βάση την υφή και με τη βοήθεια αυτής της μεθόδου μπορούμε να αποφύγουμε τη εφαρμογή κάποιων φίλτρων στην αρχική εικόνα και τον έλεγχο αν οι αποκρίσεις σε αυτά τα φίλτρα είναι καλές (πράγμα το οποίο είναι αρκετά χρονοβόρο).

2.3. Ανάλυση σε Πρωτεύουσες Συνιστώσες

Η μέθοδος της «Ανάλυσης σε Πρωτεύουσες Συνιστώσες (*Principal Components Analysis-PCA*)» είναι μια ευρέως διαδεδομένη τεχνική που χρησιμοποιείται για εφαρμογές όπως η μείωση διάστασης, η εξαγωγή χαρακτηριστικών και η οπτικοποίηση δεδομένων. Μπορεί να οριστεί ως η ορθογώνια προβολή των δεδομένων σε χώρο μικρότερης διάστασης – γνωστός ως πρωτεύων υποχώρος- έτσι ώστε η διακύμανση των δεδομένων που προβάλλονται να μεγιστοποιείται. Αλλιώς, μπορεί να οριστεί ως η γραμμική προβολή που ελαχιστοποιεί το μέσο κόστος



προβολής, γνωστό ως μέση τετραγωνική απόσταση μεταξύ των δεδομένων και των προβολών τους.

Η μέθοδος λειτουργεί ως εξής:

Έστω ένα σύνολο παρατηρήσεων $\{x_n\}$ όπου $n=1, \dots, N$ και x_n μια ευκλείδια μεταβλητή διάστασης D . Στόχος μας είναι να προβάλλουμε τα δεδομένα σε ένα διάστημα διάστασης $M < D$ καθώς μεγιστοποιούμε τη διακύμανση των προβαλλόμενων δεδομένων. Θεωρούμε ότι η τιμή του M είναι δεδομένη.

Αρχικά, θεωρούμε την προβολή σε διάστημα μίας διάστασης ($M = 1$). Μπορούμε να ορίσουμε την κατεύθυνση αυτού του χώρου χρησιμοποιώντας ένα D -διάστατο διάνυσμα u_1 , το οποίο επιλέγουμε να είναι ένα μοναδιαίο διάνυσμα έτσι ώστε $u_1^T u_1 = 1$ (σημειώνουμε ότι ενδιαφερόμαστε για την διεύθυνση που ορίζεται από το u_1 και όχι για το μέτρο του). Κάθε σημείο x_n στη συνέχεια προβάλλεται σε μια μονοδιάστατη τιμή, την $u_1^T x_n$. Το μέσο των προβαλλόμενων δεδομένων είναι $u_1^T \bar{x}$, όπου \bar{x} είναι το μέσο του συνόλου μας που δίνεται από τον τύπο (2.11):

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n \quad (2.11)$$

και η διακύμανση των προβαλλόμενων δεδομένων δίνεται από τον τύπο (2.12):

$$\frac{1}{N} \sum_{n=1}^N \{u_1^T x_n - u_1^T \bar{x}\}^2 = u_1^T S u_1 \quad (2.12)$$

όπου S αποτελείται από τα δεδομένα του πίνακα συμμεταβλητότητας, ο οποίος ορίζεται από την ποσότητα (2.13):

$$S = \frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})(x_n - \bar{x})^T \quad (2.13)$$

Στη συνέχεια, μεγιστοποιούμε τη διακύμανση που προβάλαμε $u_1^T S u_1$ ως προς το u_1 . Αυτό ουσιαστικά είναι μια υπό συνθήκη μεγιστοποίηση έτσι ώστε να αποφύγουμε $\|u_1\| \rightarrow \infty$. Η καταλληλότερος περιορισμός προκύπτει από τη συνθήκη κανονικοποίησης $u_1^T u_1 = 1$. Για να εφαρμόσουμε αυτή τη συνθήκη στο πρόβλημά μας



εισάγουμε τους «πολλαπλασιαστές *Lagrange*», επομένως πρέπει να μεγιστοποιήσουμε την ποσότητα:

$$u_1^T S u_1 + \lambda_1 (1 - u_1^T u_1)$$

Εάν θέσουμε την παράγωγο ως προς u_1 ίση με το μηδέν, βλέπουμε ότι η λύση της παραπάνω ποσότητας προκύπτει όταν:

$$S u_1 = \lambda_1 u_1,$$

από το οποίο προκύπτει ότι το u_1 πρέπει να είναι ιδιοδιάνυσμα του S . Εάν στη συνέχεια πολλαπλασιάσουμε από αριστερά με u_1^T χρησιμοποιήσουμε το ότι $u_1^T u_1 = 1$, βλέπουμε ότι η διακύμανση δίνεται από τον τύπο:

$$u_1^T S u_1 = \lambda_1.$$

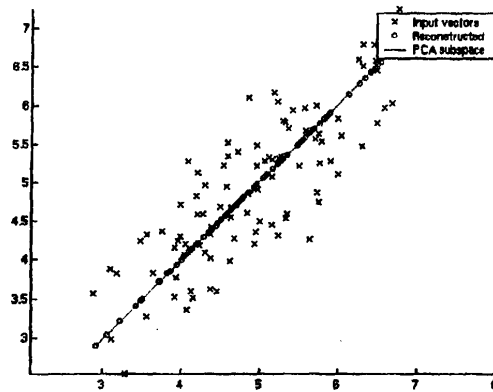
Επομένως, η διακύμανση μεγιστοποιείται όταν θέσουμε το u_1 ίσο με το ιδιοδιάνυσμα που έχει τη μέγιστη ιδιοτιμή λ_1 . Αυτό το ιδιοδιάνυσμα είναι γνωστό ως *Πρώτη Πρωτεύουσα Συνιστώσα*. Μπορούμε να ορίσουμε επιπλέον πρωτεύουσες συνιστώσες επιλέγοντας κάθε νέα διεύθυνση να είναι εκείνη που μεγιστοποιεί την προβαλλόμενη διακύμανση ανάμεσα σε όλες τις πιθανές κατευθύνσεις, οι οποίες είναι ορθογώνιες σε αυτές που ήδη υπολογίστηκαν.

Εάν θεωρήσουμε τη γενική περίπτωση ενός M -διάστατου διαστήματος προβολής, η βέλτιστη γραμμική προβολή για την οποία η διακύμανση των προβαλλόμενων δεδομένων μεγιστοποιείται, ορίζεται από τα M ιδιοδιανύσματα u_1, \dots, u_M των δεδομένων του πίνακα συμμεταβλητότητας S που αντιστοιχούν στις M μεγαλύτερες ιδιοτιμές $\lambda_1, \dots, \lambda_M$.

Συνολικά, η Ανάλυση σε Πρωτεύουσες Συνιστώσες συμπεριλαμβάνει την εκτίμηση του μέσου \bar{x} και του πίνακα συμμεταβλητότητας S του συνόλου δεδομένων και στη συνέχεια την εύρεση των M ιδιοδιανυσμάτων του S που αντιστοιχούν στις M μεγαλύτερες ιδιοτιμές.

Στο Σχήμα 2.10 φαίνεται ένα παράδειγμα της Ανάλυσης σε Πρωτεύουσες συνιστώσες.





Σχήμα 2.10: Παράδειγμα εύρεσης του άξονα μέγιστης μεταβλητότητας με τη μέθοδο της Ανάλυσης σε Πρωτεύουσες Συνιστώσες

Η μείωση της διάστασης λύνει ένα σημαντικό πρόβλημα. Το πρόβλημα αυτό αναφέρεται ως «κατάρτα της διάστασης» και είναι το φαινόμενο σύμφωνα με το οποίο καθώς αυξάνει η διάσταση, τα δεδομένα γίνονται πιο αραιά στο χώρο που καταλαμβάνουν. Επομένως όταν έχουμε δεδομένα μεγάλης διάστασης, η επίδοση των τεχνικών ομαδοποίησης δεν είναι καλή. Μία καλή μέθοδος για να αντιμετωπιστεί το πρόβλημα αυτό είναι να εφαρμόσουμε τεχνικές μείωσης της διάστασης.

Η μείωση της διάστασης έχει και περαιτέρω σημαντικά πλεονεκτήματα. Αρχικά, μπορεί να εξαλείψει μη σχετικά χαρακτηριστικά των δεδομένων και να μειώσει, κατά ένα μέρος, το θόρυβο. Επιπλέον, μπορεί να οδηγήσει σε ένα πιο κατανοητό μοντέλο καθώς, πλέον, το μοντέλο θα περιέχει λιγότερα χαρακτηριστικά.

2.4. Ομαδοποίηση Δεδομένων με τον Αλγόριθμο K-μέσων

Η διαδικασία της ομαδοποίησης περιλαμβάνει τα ακόλουθα βήματα:

- Αναπαράσταση των χαρακτηριστικών. Στο στάδιο αυτό περιλαμβάνεται και η εξαγωγή (*feature extraction*) και/ή επιλογή χαρακτηριστικών (*feature selection*).
- Επιλογή ενός μέτρου ομοιότητας κατάλληλου για το πεδίο ορισμού των δεδομένων.

- Ομαδοποίηση.
- Αφαίρεση δεδομένων (data abstraction).
- Αξιολόγηση του αποτελέσματος.

Με τον όρο αναπαράσταση των χαρακτηριστικών αναφερόμαστε στο πλήθος των ομάδων και στο πλήθος των χαρακτηριστικών τα οποία είναι διαθέσιμα σε έναν αλγόριθμο ομαδοποίησης προς αξιοποίηση. Η επιλογή των χαρακτηριστικών είναι η διαδικασία της αναγνώρισης του πιο σημαντικού υποσυνόλου χαρακτηριστικών των δεδομένων που μπορούν να χρησιμοποιηθούν κατά την ομαδοποίηση. Η εξαγωγή χαρακτηριστικών είναι η χρήση ενός ή περισσότερων μετασχηματισμών χαρακτηριστικών εισόδου προκειμένου να παραχθούν νέα salient χαρακτηριστικά. Οι τεχνικές αυτές χρησιμοποιούνται για την εύρεση ενός κατάλληλου συνόλου χαρακτηριστικών τα οποία μπορούν να χρησιμοποιηθούν κατά την ομαδοποίηση.

Η επιλογή του μέτρου ομοιότητας καθορίζει τον τρόπο υπολογισμού της ομοιότητας δύο σημείων.

Η ομαδοποίηση, δηλαδή η διαίρεση του συνόλου σε ομάδες όμοιων αντικειμένων, μπορεί να γίνει με πολλούς τρόπους. Ο τρόπος ο οποίος ακολουθείται εξαρτάται από τον αλγόριθμο ομαδοποίησης που χρησιμοποιείται σε κάθε περίπτωση.

Αφαίρεση δεδομένων είναι η διαδικασία εξαγωγής μιας απλής και συμπαγούς αναπαράστασης του συνόλου δεδομένων.

Η αξιολόγηση του αποτελέσματος έγκειται στην εκτίμηση του αποτελέσματος της ομαδοποίησης, δηλαδή στο κατά πόσο η ομαδοποίηση που προκύπτει είναι καλή ή όχι.

Ο πιο ευρέως διαδεδομένος και απλός αλγόριθμος τετραγωνικού σφάλματος είναι ο αλγόριθμος K-μέσων. Ο αλγόριθμος ξεκινά με μια τυχαία διαμέριση των εικονοστοιχείων σε ομάδες και στη συνέχεια, επανατοποθετεί τα εικονοστοιχεία στις ομάδες με βάση την ομοιότητα μεταξύ αυτών και των κέντρων των ομάδων. Έπειτα,

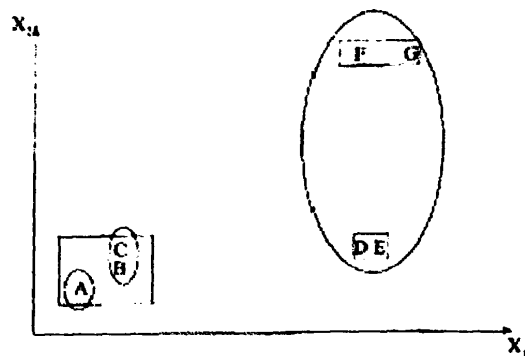


τα κέντρα κάθε ομάδας ενημερώνονται με βάση τα σημεία που ανατίθενται στη συγκεκριμένη ομάδα. Επαναλαμβάνουμε τις αναθέσεις και τις ενημερώσεις μέχρι να παραμείνει αμετάβλητη η ομάδα καθενός σημείου ή μέχρι το τετραγωνικό σφάλμα να μειωθεί σημαντικά. Παρακάτω φαίνονται τα βήματα του αλγορίθμου K -μέσων:

- Επιλέγουμε K κέντρα έτσι ώστε να συμπίπτουν με K τυχαία επιλεγμένα εικονοστοιχεία.
- Επαναλαμβάνουμε
 - Δημιουργούμε K ομάδες αναθέτοντας κάθε εικονοστοιχείο στο κοντινότερο του κέντρο.
 - Ενημερώνουμε το κέντρο της κάθε ομάδας.
- Μέχρι τα κέντρα να μη μεταβάλλονται ή την ελαχιστοποίηση του τετραγωνικού σφάλματος.

Ο αλγόριθμος K -μέσων είναι ιδιαίτερα διαδεδομένος καθώς έχει χρονική πολυπλοκότητα $O(n)$ – όπου n ο αριθμός των εικονοστοιχείων. Ένα βασικό μειονέκτημα του αλγορίθμου αυτού είναι ότι το αποτέλεσμα της εφαρμογής του εξαρτάται από την αρχική διαμέριση. Επομένως, αν η αρχική διαμέριση δεν έχει επιλεγεί σωστά μπορεί ο αλγόριθμος να συγκλίνει σε τοπικό ελάχιστο. Στο Σχήμα 2.11 μπορούμε να διακρίνουμε τη σημασία της αρχικής διαμέρισης για τον αλγόριθμο K -μέσων. Συγκεκριμένα στο παράδειγμα υπάρχουν 7 ομάδες. Αν ξεκινήσουμε την εφαρμογή του K -μέσων έχοντας τα πρότυπα A , B και C ως αρχικά κέντρα γύρω από τα οποία θα δημιουργηθούν οι τρεις ομάδες, η τελική μας διαμέριση θα είναι $\{\{A\}, \{B, C\}, \{D, E, F, G\}\}$. Όμως, στην παραπάνω περίπτωση, το τετραγωνικό σφάλμα είναι πολύ μεγαλύτερο σε σχέση με αυτό που έχουμε όταν η τελική διαμέριση είναι η βέλτιστη δηλαδή, $\{\{A, B, C\}, \{D, E\}, \{F, G\}\}$. Η σωστή τελική διαμέριση παρέχεται όταν για παράδειγμα επιλεγούν ως αρχικά κέντρα των ομάδων τα πρότυπα A , D και F .





Σχήμα 2.11: Η εξάρτηση του K -μέσων από την αρχική διαμέριση

2.5. Πειραματικά Αποτελέσματα

Στο Σχήμα 2.12 φαίνεται μία εικόνα της βάσης Berkeley, της οποίας τα εικονοστοιχεία θα ομαδοποιήσουμε σε 5, 10 και 15 ομάδες με τη χρήση του αλγορίθμου K -μέσων στους «αντιπροσώπους» των ομάδων. Η αρχική επιλογή των κέντρων γίνεται τυχαία και στη συνέχεια υπολογίζουμε τους «αντιπροσώπους» (τα μέσα ή τα medoid) κάθε υπερ-εικονοστοιχείου, τα οποία ομαδοποιούμε με τον αλγόριθμο K -μέσων.

Το medoid υπολογίζεται ως εξής: Παίρνουμε κάθε εικονοστοιχείο του υπερ-εικονοστοιχείου και αθροίζουμε τις αποστάσεις που απέχει από όλα τα υπόλοιπα. Το εικονοστοιχείο που έχει το μικρότερο άθροισμα αποστάσεων είναι το medoid.

Στο Σχήμα 2.12 α) βλέπουμε την αρχική εικόνα ενώ στο Σχήμα 2.12 β) βλέπουμε τα υπερ-εικονοστοιχεία για την εικόνα αυτή.

Στα σχήματα 2.13, 2.14, 2.15 φαίνονται τα όρια περιοχών που προκύπτουν ύστερα από την εφαρμογή του αλγορίθμου K -μέσων στα μέσα των υπερ-εικονοστοιχείων και με τη χρήση των RGB -συντεταγμένων στην αρχική εικόνα του σχήματος 2.12 για 5, 10 και 15 κέντρα, αντίστοιχα.



α)

β)

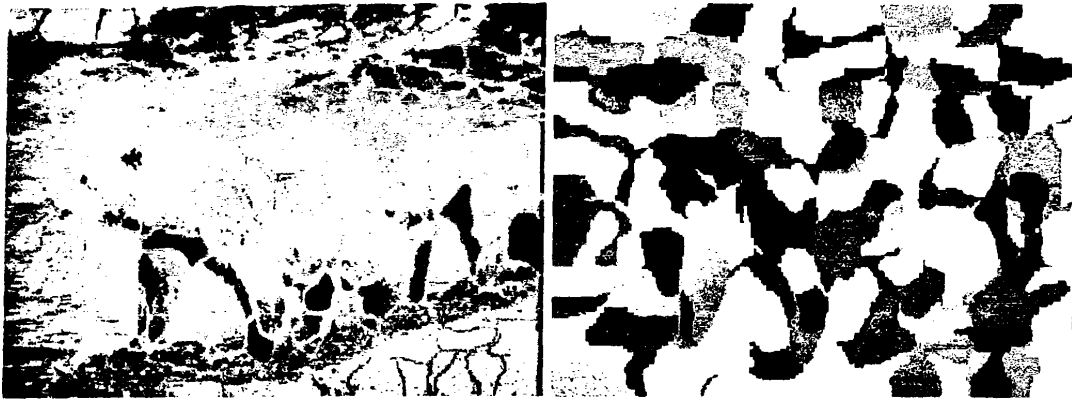
Σχήμα 2.12: α) Αρχική εικόνα, β) Υπερεικονοστοιχεία



Σχήμα 2.13: Όρια περιοχών που προκύπτουν για $K=5$ κέντρα με τη χρήση των RGB-συντεταγμένων και ύστερα από ομαδοποίηση των μέσων των υπερ-εικονοστοιχείων

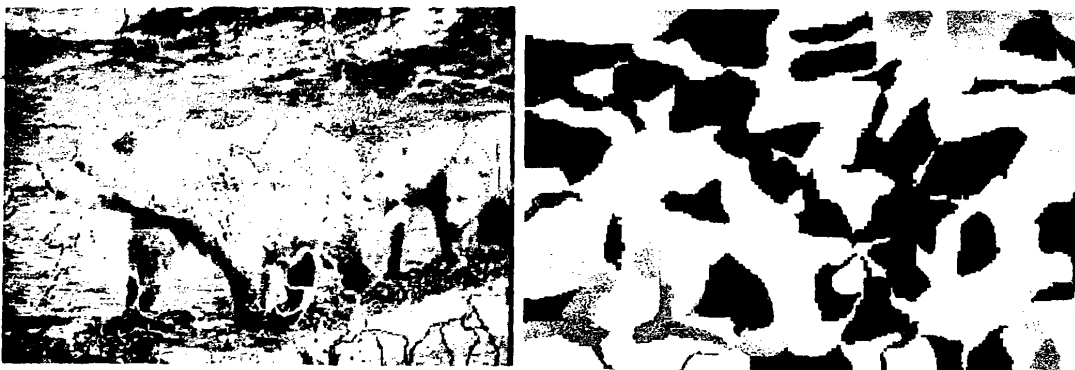


Σχήμα 2.14: Όρια περιοχών που προκύπτουν για $K=10$ κέντρα με τη χρήση των RGB-συντεταγμένων και ύστερα από ομαδοποίηση των μέσων των υπερ-εικονοστοιχείων

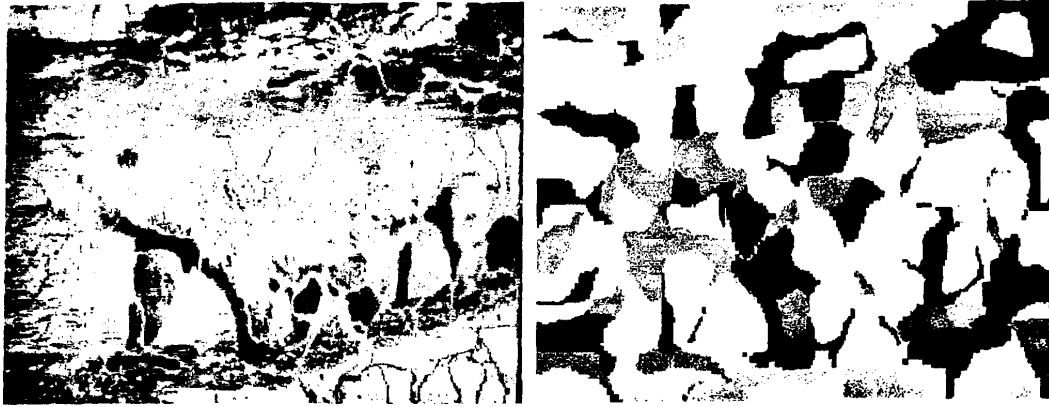


Σχήμα 2.15: Όρια περιοχών που προκύπτουν για $K=15$ κέντρα με τη χρήση των RGB -συντεταγμένων και ύστερα από ομαδοποίηση των μέσων των υπερ-εικονοστοιχείων

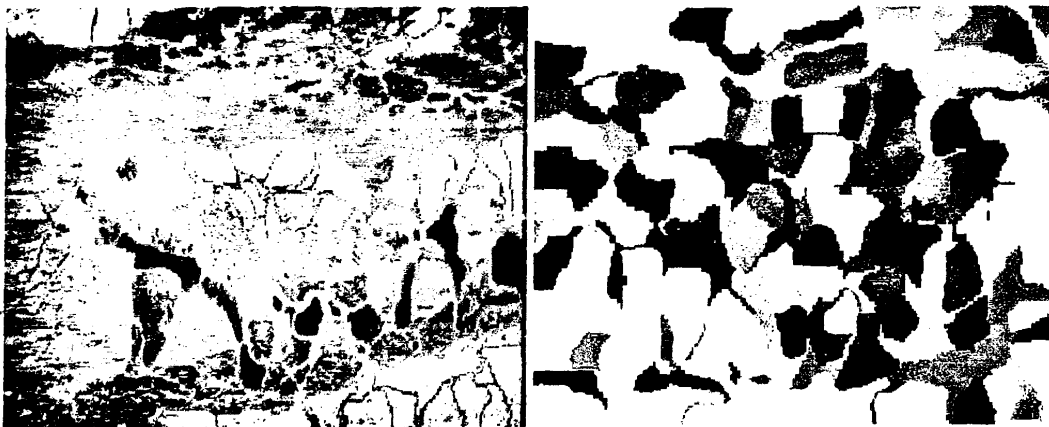
Στα σχήματα 2.16, 2.17, 2.18 φαίνονται τα όρια περιοχών που προκύπτουν ύστερα από την εφαρμογή του αλγορίθμου K -μέσων στα μέσα των υπερ-εικονοστοιχείων και με τη χρήση των MRF -χαρακτηριστικών στην αρχική εικόνα του σχήματος 2.12 για 5, 10 και 15 κέντρα, αντίστοιχα.



Σχήμα 2.16: Όρια περιοχών που προκύπτουν για $K=5$ κέντρα με τη χρήση των χαρακτηριστικών MRF και ύστερα από ομαδοποίηση των μέσων των υπερ-εικονοστοιχείων



Σχήμα 2.17: Όρια περιοχών που προκύπτουν για $K=10$ κέντρα με τη χρήση των χαρακτηριστικών MRF και ύστερα από ομαδοποίηση των μέσων των υπερ-εικονοστοιχείων



Σχήμα 2.18: Όρια περιοχών που προκύπτουν για $K=15$ κέντρα με τη χρήση των χαρακτηριστικών MRF και ύστερα από ομαδοποίηση των μέσων των υπερ-εικονοστοιχείων

Στα σχήματα 2.19, 2.20, 2.21 φαίνονται τα όρια περιοχών που προκύπτουν ύστερο από την εφαρμογή του αλγορίθμου K -μέσων στα medoids των υπερ-εικονοστοιχείων και με τη χρήση των RGB -συντεταγμένων στην αρχική εικόνα του σχήματος 2.12 για 5, 10 και 15 κέντρα, αντίστοιχα.



Σχήμα 2.19: Όρια περιοχών που προκύπτουν για $K=5$ κέντρα με τη χρήση των RGB συντεταγμένων και ύστερα από ομαδοποίηση των medoid των υπερ-εικονοστοιχείων

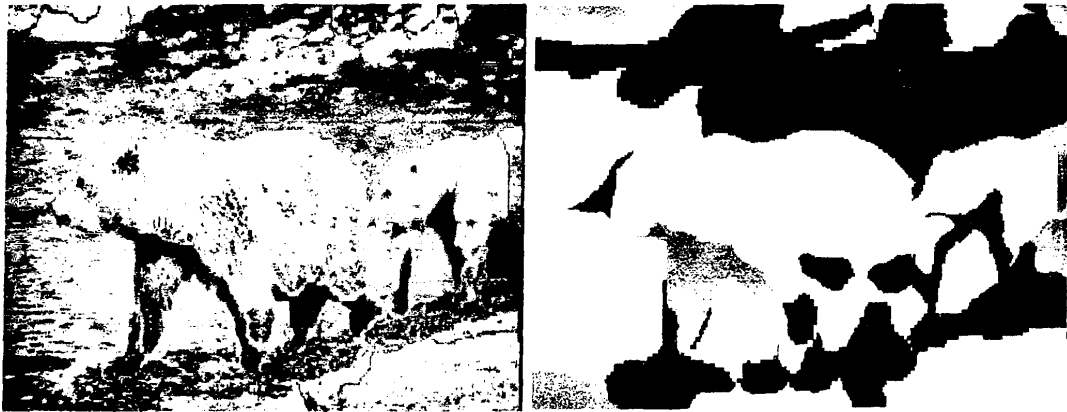


Σχήμα 2.20: Όρια περιοχών που προκύπτουν για $K=10$ κέντρα με τη χρήση των RGB -συντεταγμένων και ύστερα από ομαδοποίηση των medoid των υπερ-εικονοστοιχείων

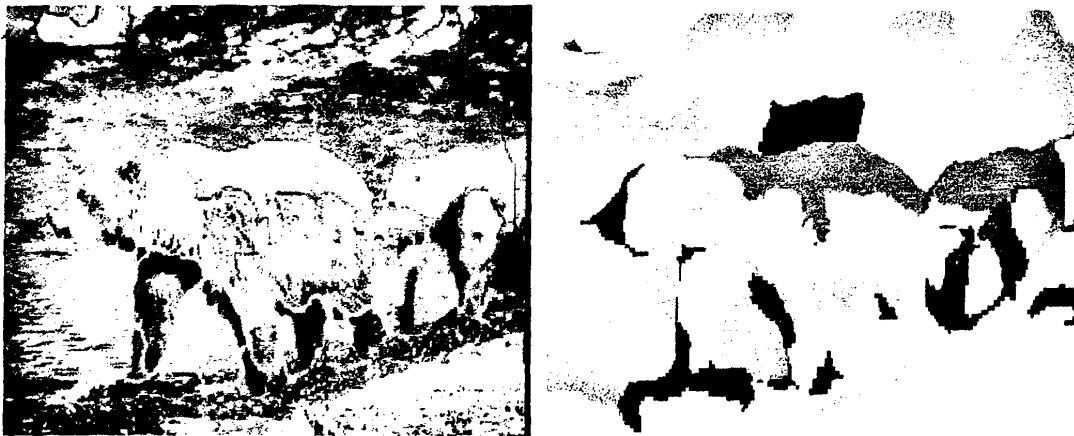


Σχήμα 2.21: Όρια περιοχών που προκύπτουν για $K=15$ κέντρα με τη χρήση των RGB -συντεταγμένων και ύστερα από ομαδοποίηση των medoid των υπερ-εικονοστοιχείων

Στα σχήματα 2.22, 2.23, 2.24 φαίνονται τα όρια περιοχών που προκύπτουν ύστερα από την εφαρμογή του αλγορίθμου K -μέσων στα medoids των υπερ-εικονοστοιχείων και με τη χρήση των χαρακτηριστικών MRF στην αρχική εικόνα του σχήματος 2.12 για 5, 10 και 15 κέντρα, αντίστοιχα.



Σχήμα 2.22: Όρια περιοχών που προκύπτουν για $K=5$ κέντρα με τη χρήση των χαρακτηριστικών MRF και ύστερα από ομαδοποίηση των medoid των υπερ-εικονοστοιχείων



Σχήμα 2.23: Όρια περιοχών που προκύπτουν για $K=10$ κέντρα με τη χρήση των χαρακτηριστικών MRF και ύστερα από ομαδοποίηση των medoid των υπερ-εικονοστοιχείων



Σχήμα 2.24: Όρια περιοχών που προκύπτουν για $K=15$ κέντρα με τη χρήση των χαρακτηριστικών MRF και ύστερα από ομαδοποίηση των medoid των υπερεικονοστοιχείων

2.5.1. Δείκτες Αξιολόγησης της Κατάμησης

Για να ελέγξουμε πόσο «καλά» είναι τα αποτελέσματα που προέκυψαν από τον αλγόριθμό μας, τα συγκρίνουμε με τις καταταμήσεις που έχουν γίνει από τους ανθρώπους, με βάση κάποιους δείκτες. Οι δείκτες που χρησιμοποιήσαμε στην εργασία αυτή, είναι οι αυτές που αναφέρονται παρακάτω.

Probabilistic Rand Index (RI) :Για κάθε ζεύγος εικονοστοιχείων x_i και x_j , ελέγχει αν οι ετικέτες που προέκυψαν από τον αλγόριθμό μας, είναι ίδιες με αυτές των ανθρώπινων καταταμήσεων. Συγκεκριμένα, παίρνει ανά δύο τα εικονοστοιχεία της κατάμησης και ελέγχει αν ανήκουν στην ίδια ή σε διαφορετικές ομάδες. Έτσι, αν $X = \{x_1, x_2, \dots, x_n\}$ η κατάταμήσή μας και $Y = \{y_1, y_2, \dots, y_n\}$ οι καταταμήσεις των εθελοντών, τότε ορίζεται η ποσότητα:

$$RI = \frac{a+b}{a+b+c+d} \quad (2.14)$$

όπου:

a : το πλήθος των ζευγών εικονοστοιχείων που ανήκουν στην ίδια ομάδα στο X και στην ίδια στο Y ,

b : το πλήθος των ζευγών που ανήκουν σε διαφορετική ομάδα στο X και σε διαφορετική στο Y ,

c : το πλήθος των ζευγών που ανήκουν στην ίδια ομάδα στο X , αλλά σε διαφορετικές ομάδες στο Y και

d : το πλήθος των ζευγών που ανήκουν σε διαφορετικές ομάδες στο X , αλλά στην ίδια ομάδα στο Y .

Το RI παίρνει τιμές στο διάστημα $[0, 1]$. Όσο μεγαλύτερη είναι η τιμή του, τόσο περισσότερο οι καταταμήσεις μοιάζουν και επομένως το αποτέλεσμα είναι καλύτερο. Τιμές ίσες με το μηδέν, δείχνουν ότι όλες οι ανθρώπινες καταταμήσεις είναι διαφορετικές από αυτές του αλγορίθμου μας, ενώ τιμές ίσες με το ένα, δείχνουν ότι όλες οι καταταμήσεις είναι ίδιες, επομένως και το αποτέλεσμα της κατάταμής μας «καλό» [6].

Variation of Information (Vol): Ορίζει την απόσταση μεταξύ δύο καταταμήσεων ως προς τη διαφορά πληροφορίας, με βάση τον τύπο:

$$VI(X;Y) = H(X) + H(Y) - 2I(X,Y) \quad (2.15)$$

Η απόσταση αυτή είναι η μέση υπό συνθήκη εντροπία της κατάταμησης που προέκυψε από τον αλγόριθμό μας, δοθέντων των ανθρώπινων καταταμήσεων, και έτσι μετριέται το ποσό τυχαιότητας της κατάταμής μας. Για τον υπολογισμό της ποσότητας αυτής, υπολογίζονται οι εντροπίες (αποστάσεις) και η αμοιβαία πληροφορία μεταξύ των εικονοστοιχείων με τις ίδιες ετικέτες της κατάταμής μας με τις ανθρώπινες καταταμήσεις. Όσο μικρότερη είναι η απόσταση αυτή, τόσο καλύτερη σημαίνει ότι είναι η κατάταμής μας [5].

Boundary Displacement Error (BDE): Κατασκευάζει χάρτες περιοχών υπολογίζοντας τις παραγώγους για καθεμία από τις δύο εικόνες και υπολογίζει την απόσταση μεταξύ των ορίων της κατάταμησης που προέκυψε από τον αλγόριθμό μας, με τις ανθρώπινες καταταμήσεις και υπολογίζει την προαναφερθείσα ποσότητα. Όσο μικρότερη τιμή έχει η ποσότητα αυτή, τόσο καλύτερο είναι το αποτέλεσμα της κατάταμησης που έχουμε [1].



Global Consistency Error (GCE): Μετρά την έκταση στην οποία μια κατάτμηση μπορεί να ιδωθεί ως εκλέπτυνση μιας άλλης. Παίρνει τιμές στο διάστημα $[0, 1]$ και όσο μικρότερη τιμή έχει, τόσο καλύτερη κατάτμηση επιτυγχάνεται [3].

Στον Πίνακα 2.1 φαίνονται τα αποτελέσματα της κατάτμησης στις εικόνες της βάσης του Berkeley με τη χρήση των *MRF* χαρακτηριστικών και ομαδοποίηση των μέσων διανυσμάτων των υπερεικονοστοιχείων σε $K = 5$ περιοχές.

Στον Πίνακα 2.2 φαίνονται τα αποτελέσματα της κατάτμησης στις εικόνες της βάσης του Berkeley με τη χρήση των *MRF* χαρακτηριστικών και ομαδοποίηση των μέσων διανυσμάτων των υπερεικονοστοιχείων σε $K = 10$ περιοχές.

Στον Πίνακα 2.3 φαίνονται τα αποτελέσματα της κατάτμησης στις εικόνες της βάσης του Berkeley με τη χρήση των *MRF* χαρακτηριστικών και ομαδοποίηση των μέσων διανυσμάτων των υπερεικονοστοιχείων σε $K = 15$ περιοχές.

Πίνακας 2.1: Αποτελέσματα κατάτμησης εικόνων με χρήση *MRF*-χαρακτηριστικών και ύστερα ομαδοποίηση των μέσων διανυσμάτων των υπερεικονοστοιχείων σε $K = 5$ περιοχές

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.515	0.021	0.514
VoI	3.520	0.268	3.558
BDE	21.304	4.290	18.880
GCE	0.432	0.043	0.455



Πίνακας 2.2: Αποτελέσματα κατάτμησης εικόνων με χρήση *MRF*-χαρακτηριστικών και ύστερα ομαδοποίηση των μέσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 10$ περιοχές

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.626	0.047	0.651
VoI	4.442	0.229	4.512
BDE	17.411	3.916	15.134
GCE	0.606	0.069	0.661

Πίνακας 2.3: Αποτελέσματα κατάτμησης εικόνων με χρήση *MRF*-χαρακτηριστικών και ύστερα ομαδοποίηση των μέσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 15$ περιοχές

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.665	0.067	0.708
VoI	4.996	0.172	5.071
BDE	16.514	3.695	14.626
GCE	0.618	0.072	0.674

Από τα πειραματικά αποτελέσματα των πινάκων 2.1, 2.2 και 2.3 παρατηρούμε ότι με τη χρήση του αλγορίθμου K -μέσων και με τη χρήση των *MRF*-χαρακτηριστικών, όσο το πλήθος των κέντρων αυξάνεται, τόσο η τιμή του Rand Index και του Boundary Detection Error βελτιώνεται.

Στον Πίνακα 2.4 φαίνονται τα αποτελέσματα της κατάτμησης στις εικόνες της βάσης του Berkeley με τη χρήση των *MRF* χαρακτηριστικών και ομαδοποίηση των ενδιάμεσων (medoids) διανυσμάτων των υπερεικονοστοιχείων σε $K = 5$ περιοχές.



Πίνακας 2.4: Αποτελέσματα κατάτμησης εικόνων με χρήση *MRF*-χαρακτηριστικών και ύστερα ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 5$ περιοχές

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.707	0.049	0.735
VoI	2.695	0.238	2.521
BDE	14.575	4.118	11.908
GCE	0.333	0.040	0.343

Στον Πίνακα 2.5 φαίνονται τα αποτελέσματα της κατάτμησης στις εικόνες της βάσης του Berkeley με τη χρήση των *MRF* χαρακτηριστικών και ομαδοποίηση των ενδιάμεσων (medoids) διανυσμάτων των υπερεικονοστοιχείων σε $K = 10$ περιοχές.

Πίνακας 2.5: Αποτελέσματα κατάτμησης εικόνων με χρήση *MRF*-χαρακτηριστικών και ύστερα ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 10$ περιοχές

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.725	0.068	0.772
VoI	3.051	0.205	3.000
BDE	14.157	4.134	11.665
GCE	0.380	0.047	0.394

Στον Πίνακα 2.6 φαίνονται τα αποτελέσματα της κατάτμησης στις εικόνες της βάσης του Berkeley με τη χρήση των *MRF* χαρακτηριστικών και ομαδοποίηση των ενδιάμεσων (medoids) διανυσμάτων των υπερεικονοστοιχείων σε $K = 15$ περιοχές.



Πίνακας 2.6: Αποτελέσματα κατάτμησης εικόνων με χρήση *MRF*-χαρακτηριστικών και ύστερα ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 15$ περιοχές

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.725	0.075	0.773
VoI	3.301	0.193	3.261
BDE	14.573	4.133	12.108
GCE	0.377	0.048	0.397

Με τη χρήση του αλγορίθμου K -μέσων στα ενδιάμεσα διανύσματα και των *MRF*-χαρακτηριστικών, παρατηρούμε ότι όσο το πλήθος των κέντρων αυξάνεται, τόσο καλύτερο Rand Index παίρνουμε. Επιπλέον, τα αποτελέσματα της ομαδοποίησης των medoids σε σχέση με αυτά της ομαδοποίησης των μέσων, με τη χρήση των *MRF*-χαρακτηριστικών, είναι καλύτερα. Αυτό γίνεται κατανοητό, αν παρατηρήσουμε ότι τιμή του Rand Index είναι αρκετά μεγαλύτερη στην περίπτωση της ομαδοποίησης των medoids.

Στον Πίνακα 2.7 φαίνονται τα αποτελέσματα της κατάτμησης στις εικόνες της βάσης του Berkeley με τη χρήση των *RGB* συντεταγμένων και ομαδοποίηση των μέσων διανυσμάτων των υπερεικονοστοιχείων σε $K = 5$ περιοχές.

Στον Πίνακα 2.8 φαίνονται τα αποτελέσματα της κατάτμησης στις εικόνες της βάσης του Berkeley με τη χρήση των *RGB* συντεταγμένων και ομαδοποίηση των μέσων διανυσμάτων των υπερεικονοστοιχείων σε $K = 10$ περιοχές.

Πίνακας 2.7: Αποτελέσματα κατάτμησης εικόνων με χρήση *RGB* συντεταγμένων και ύστερα ομαδοποίηση των μέσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 5$ περιοχές

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.600	0.043	0.624
VoI	4.029	0.256	4.072
BDE	16.921	4.677	14.672
GCE	0.568	0.069	0.627

Πίνακας 2.8: Αποτελέσματα κατάτμησης εικόνων με χρήση *RGB* συντεταγμένων και ύστερα ομαδοποίηση των μέσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 10$ περιοχές

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.648	0.062	0.685
VoI	4.626	0.223	4.691
BDE	16.602	4.588	14.474
GCE	0.606	0.073	0.669

Στον Πίνακα 2.9 φαίνονται τα αποτελέσματα της κατάτμησης στις εικόνες της βάσης του Berkeley με τη χρήση των *RGB* συντεταγμένων και ομαδοποίηση των μέσων διανυσμάτων των υπερεικονοστοιχείων σε $K = 15$ περιοχές.



Πίνακας 2.9: Αποτελέσματα κατάτμησης εικόνων με χρήση *RGB* συντεταγμένων και ύστερα ομαδοποίηση των μέσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 15$ περιοχές

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.665	0.070	0.707
VoI	4.923	0.207	5.002
BDE	16.598	4.614	14.406
GCE	0.598	0.073	0.645

Με τη χρήση του αλγορίθμου K -μέσων και των *RGB*-συντεταγμένων, παρατηρούμε ότι όσο το πλήθος των κέντρων αυξάνεται, τόσο καλύτερο Rand Index και ελαφρώς καλύτερο Boundary Detextion Error παίρνουμε. Τα αποτελέσματα είναι λίγο καλύτερα σε σχέση με τη χρήση των *MRF*-χαρακτηριστικών.

Στον Πίνακα 2.10 φαίνονται τα αποτελέσματα της κατάτμησης στις εικόνες της βάσης του Berkeley με τη χρήση των *RGB* συντεταγμένων και ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερεικονοστοιχείων σε $K = 5$ περιοχές.

Πίνακας 2.10: Αποτελέσματα κατάτμησης εικόνων με χρήση *RGB* συντεταγμένων και ύστερα ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 5$ περιοχές

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.702	0.050	0.735
VoI	2.922	0.243	2.883
BDE	13.972	4.027	11.954
GCE	0.385	0.043	0.406



Στον Πίνακα 2.11 φαίνονται τα αποτελέσματα της κατάτμησης στις εικόνες της βάσης του Berkeley με τη χρήση των *RGB* συντεταγμένων και ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερεικονοστοιχείων σε $K = 10$ περιοχές.

Πίνακας 2.11: Αποτελέσματα κατάτμησης εικόνων με χρήση *RGB* συντεταγμένων και ύστερα ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 10$ περιοχές

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.715	0.072	0.760
VoI	3.365	0.207	3.295
BDE	14.504	3.933	12.223
GCE	0.405	0.047	0.429

Στον Πίνακα 2.12 φαίνονται τα αποτελέσματα της κατάτμησης στις εικόνες της βάσης του Berkeley με τη χρήση των *RGB* συντεταγμένων και ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερεικονοστοιχείων σε $K = 15$ περιοχές.

Πίνακας 2.12: Αποτελέσματα κατάτμησης εικόνων με χρήση *RGB* συντεταγμένων και ύστερα ομαδοποίηση των ενδιάμεσων διανυσμάτων των υπερ-εικονοστοιχείων σε $K = 15$ περιοχές

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.716	0.075	0.766
VoI	3.633	0.195	3.628
BDE	14.922	3.896	12.652
GCE	0.390	0.052	0.397



Με τη χρήση του αλγορίθμου *K*-μέσων για την ομαδοποίηση των ενδιάμεσων διανυσμάτων και των *RGB*-συντεταγμένων, παρατηρούμε ότι όσο το πλήθος των κέντρων αυξάνεται, τόσο καλύτερο Rand Index παίρνουμε.

Συνολικά, παρατηρούμε ότι με την εφαρμογή του αλγορίθμου *K*-μέσων για την ομαδοποίηση των ενδιάμεσων και την χρήση των MRF χαρακτηριστικών, παίρνουμε καλύτερα αποτελέσματα.



ΚΕΦΑΛΑΙΟ 3. ΚΑΤΑΤΜΗΣΗ ΕΙΚΟΝΑΣ ΜΕ ΤΗ ΘΕΩΡΙΑ ΡΥΘΜΟΥ - ΠΑΡΑΜΟΡΦΩΣΗΣ

-
- 3.1 Εισαγωγή
 - 3.2 Θεωρία Ρυθμού – Παραμόρφωσης
 - 3.3 Η Κανονική Κατανομή
 - 3.4 Η Κατανομή Student's-*t*
 - 3.5 Κατάτμηση Φυσικών Εικόνων χωρίς Επίβλεψη
 - 3.6 Πειραματικά Αποτελέσματα
-

3.1. Εισαγωγή

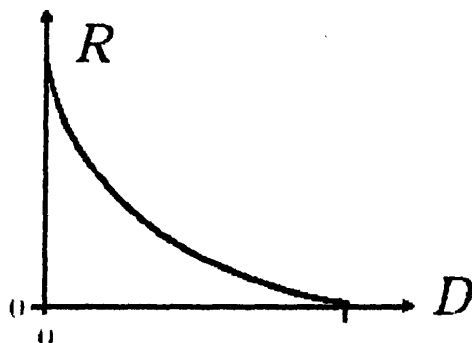
Για την κατάτμηση μιας φυσικής εικόνας, χρησιμοποιείται η θεωρία Ρυθμού – Παραμόρφωσης, με στόχο την ελαχιστοποίηση των συνολικών bits που απαιτούνται για την κατάτμηση της εικόνας. Η θεωρία αυτή, περιγράφεται στην ενότητα 3.2. Τα δεδομένα του κάθε υπερ-εικονοστοιχείου μοντελοποιούνται με τη χρήση κανονικής κατανομής ή της εύρωστης κατανομής Student's-*t*. Στις ενότητες 3.3 και 3.4 περιγράφονται κατανομές αυτές. Στην ενότητα 3.5 περιγράφεται ο αλγόριθμος που χρησιμοποιείται για την κατάτμηση εικόνων.

3.2. Θεωρία Ρυθμού – Παραμόρφωσης

Η θεωρία Ρυθμού – Παραμόρφωσης (*Rate – Distortion Theory*) [10], είναι ένας κλάδος της Θεωρίας Πληροφοριών, που έχει σχέση με το πρόβλημα του καθορισμού του ελάχιστου ποσού πληροφορίας που πρέπει να περάσει μέσα από ένα κανάλι έτσι ώστε να υπάρχει κάποια συγκεκριμένη επιτρεπόμενη παραμόρφωση της πληροφορίας ύστερα από την μεταφορά της πληροφορίας στον παραλήπτη.



Στο Σχήμα 3.1 φαίνεται η γραφική παράσταση του ρυθμού (R) ως προς την παραμόρφωση (D). Βλέπουμε ότι οι τιμές του ρυθμού είναι αντιστρόφως ανάλογες με αυτές της παραμόρφωσης. Όταν η παραμόρφωση έχει μηδενική τιμή, ο ρυθμός μεγιστοποιείται, ενώ όταν η παραμόρφωση έχει τη μέγιστη τιμή της, ο ρυθμός έχει τιμή ίση με μηδέν.



Σχήμα 3.1: Γενική μορφή της γραφικής παράστασης Ρυθμού – Παραμόρφωσης

Η Θεωρία Ρυθμού – Παραμόρφωσης αναφέρεται στην ελάχιστη αμοιβαία πληροφορία (δηλαδή την πληροφορία/ρυθμό) που μπορεί να ανεχθεί ένα κανάλι για κάποια σύμβολα εισόδου, ενώ η παραμόρφωση δε θα πρέπει να ξεπερνάει κάποιο κατώφλι D .

Έστω ότι έχουμε M πιθανά πηγαία μηνύματα που αναπαρίστανται σαν μια κωδική λέξη μήκους n . Υποθέτουμε ότι $H(M)$ είναι το μέσο πλήθος bits που μεταδίδονται για κάθε πηγαίο μήνυμα. Τότε ο ρυθμός R ορίζεται ως:

$$R = \frac{H(M)}{n} \quad (3.1)$$

όπου $H(M)$ είναι η εντροπία του πηγαίου μηνύματος, ενώ στην περίπτωση που έχουμε ισοπίθανα πηγαία μηνύματα ισχύει:

$$H(M) = \log M \quad (3.2)$$

Το μέτρο παραμόρφωσης $d(x_i, y_j)$ για ένα στοιχείο μιας λέξης ορίζεται ως το μέτρο του ρυθμού για την αναπαράσταση του πηγαίου συμβόλου x_i ως σύμβολο y_j στην έξοδο. Η μέση παραμόρφωση \bar{d} ορίζεται ως το μέσο $d(x_i, y_j)$ ως προς όλα τα πιθανά πηγαία σύμβολα και συνδυασμούς αναπαραστάσεων συμβόλων:

$$\bar{d} = \sum_{i=1}^q \sum_{j=1}^r P(x_i)P(y_j | x_i)d(x_i, y_j) \quad (3.3)$$

όπου $P(y_j | x_i)$ είναι το κανάλι ή οι πιθανότητες μεταβάσεων των πηγαίων συμβόλων στην έξοδο. Για κάθε D -αποδεκτή υπό συνθήκη πιθανότητα (όταν η μέση παραμόρφωση είναι μικρότερη ή ίση από κάποιο αποδεκτό ή συγκεκριμένο κατώφλι, D) έχουμε μια αμοιβαία πληροφορία (ή ρυθμό πληροφορίας) που δίνεται από τον τύπο:

$$I(X;Y) = \sum_{i=1}^q \sum_{j=1}^r P(x_i)P(y_j | x_i) \log \frac{P(y_j | x_i)}{P(y_j)} \quad (3.4)$$

όπου $P(y_j) = \sum_{i=1}^q P(x_i)P(y_j | x_i)$ είναι οι πιθανότητες για το κάθε σύμβολο.

Για κάποια δοθείσα μέση παραμόρφωση D και καθορισμένες κατανομές πιθανότητας $\{P(x_i) : i = 1, 2, \dots, q\}$, ο ελάχιστος ρυθμός πληροφορίας που απαιτείται, δίνεται από τη συνάρτηση ρυθμού-παραμόρφωσης:

$$R(D) = \min_{P(y_j | x_i) \in P_D} I(X;Y) \quad (3.5)$$

που προκύπτει αν βρούμε τη D -αποδεκτή υπό συνθήκη πιθανότητα που ελαχιστοποιεί την αμοιβαία πληροφορία ως προς τις συνθήκες:

$$\sum_{j=1}^r P(y_j | x_i) = 1 \text{ για } i = 1, 2, \dots, r \quad (3.6)$$



Έτσι, αναμένουμε ότι όσο μεγαλύτερη παραμόρφωση ανεχόμαστε, τόσο μικρότερη πληροφορία παίρνουμε και αντίστροφα, αυξάνοντας το ρυθμό, τόσο μικρότερη παραμόρφωση είναι πιθανό να έχουμε.

Η συνάρτηση του ρυθμού ως προς την παραμόρφωση (Σχήμα 3.1) είναι μονότονα μειούμενη και παίρνει τιμές στο διάστημα:

$$0 \leq R(D) \leq H(X),$$

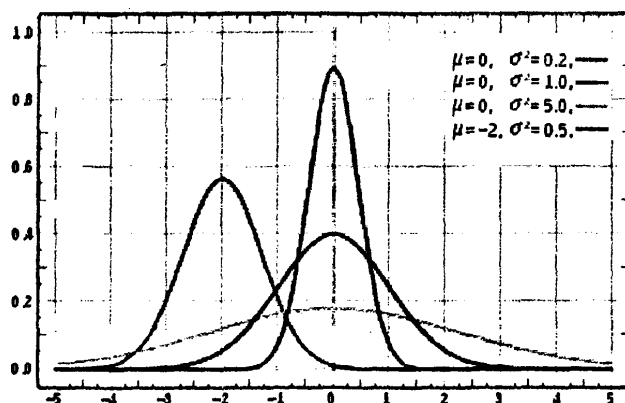
όπου $R(D_{\min}) = H(X)$ και δείχνει ότι όταν έχουμε την ελάχιστη δυνατή παραμόρφωση, έχουμε το μέγιστο ρυθμό και αντίστροφα.

3.3. Η Κανονική Κατανομή

Μια τυχαία μεταβλητή $X = x, x \in R^D$ που ακολουθεί την Κανονική κατανομή με μέση τιμή μ και πίνακα συμμεταβλητότητας Σ , δηλαδή $X \sim N(\mu, \Sigma)$ έχει συνάρτηση πυκνότητας πιθανότητας που δίνεται από τον τύπο:

$$p(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right\} \quad (3.7)$$

Η γραφική παράσταση συνάρτησης πυκνότητας πιθανότητας της Κανονικής κατανομής για μονοδιάστατα δεδομένα, είναι αυτή που φαίνεται στο Σχήμα 3.2.

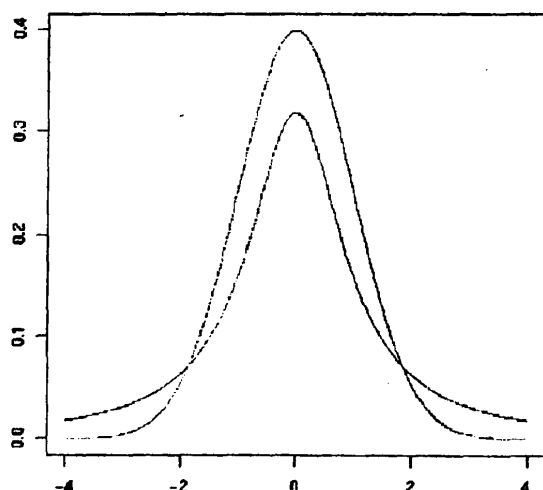


Σχήμα 3.2: Γραφική παράσταση πυκνότητας πιθανότητας Κανονικής κατανομής. Όσο μεγαλύτερη είναι η διασπορά, τόσο πιο πολύ απλώνονται τα άκρα της

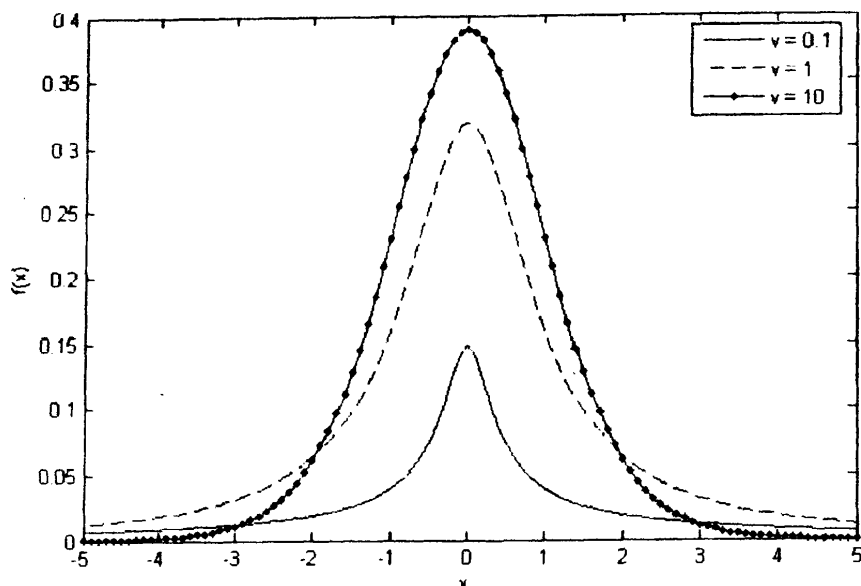
3.4. Η Κατανομή Student's-t

Το μοντέλο της Κανονικής κατανομής είναι αρκετά διαδεδομένο, έχει όμως ένα βασικό μειονέκτημα. Είναι αρκετά ευαίσθητο σε δεδομένα με θόρυβο, και μάλιστα με την παραμικρή εμφάνιση αποκλινόντων σημείων (outliers), έχει ως αποτέλεσμα τη μετατόπιση της έλλειψης που χαρακτηρίζει την ομάδα, προς τα αποκλίνοντα σημεία, μειώνοντας έτσι την αξιοπιστία του μοντέλου.

Λύση στο πρόβλημα αυτό, μπορεί ν' αποτελέσει η Student's-t κατανομή, η οποία παρουσιάζει ανθεκτικότητα στην παρουσία σημείων που αποκλίνουν από το μοντέλο. Μία εξήγηση αυτής της ευρωστίας της Student's-t κατανομής μπορεί να γίνει κατανοητή από τη γραφική παράσταση του σχήματος 3.3 των συναρτήσεων πυκνότητας πιθανότητας για μία Κανονική και μία Student's-t κατανομή. Παρατηρήστε πώς προσαρμόζονται τα άκρα της Student's-t κατανομής για να μοντελοποιήσει τα αποκλίνοντα σημεία. Επιπλέον, όπως φαίνεται στο Σχήμα 3.4 ανάλογα με την τιμή της παραμέτρου ν , η καμπύλη της γραφικής παράστασης απλώνει τα άκρα της για να μοντελοποιήσει τα αποκλίνοντα σημεία, αντί να μετακινήσει τη μέση τιμή της. Όσο μικρότερη είναι η τιμή του ν , τόσο πιο θορυβώδη είναι τα δεδομένα, άρα υπάρχουν πολλές αποκλίσεις από το γκαουσιανό μοντέλο περιγραφής τους.



Σχήμα 3.3: Γραφική παράσταση της συνάρτησης πυκνότητας πιθανότητας Κανονικής (κυανό) και Student's-t (κόκκινο)



Σχήμα 3.4: Γραφική παράσταση της συνάρτησης πυκνότητας πιθανότητας της Student's-t κατανομής για διάφορες τιμές της παραμέτρου ν . Παρατηρήστε ότι καθώς το ν μεγαλώνει, η καμπύλη προσεγγίζει αυτή της Κανονικής κατανομής

Όταν λέμε ότι μια p -διάστατη τυχαία μεταβλητή Y ακολουθεί τη Student's- t κατανομή, δηλαδή $Y \sim T(\mu, \Sigma, \nu)$, εννοούμε ότι, πρώτον, δοθέντος ενός διανύσματος βαρών τ , η Y ακολουθεί μια πολυδιάστατη κανονική κατανομή με μέση τιμή μ και πίνακα συμμεταβλητότητας Σ και δεύτερον η τυχαία μεταβλητή τ ακολουθεί την κατανομή Gamma με παράμετρο ν . Πιο συγκεκριμένα:

$$Y | \mu, \Sigma, \nu \sim N(\mu, \Sigma / \tau)$$

$\tau | \mu, \Sigma, \nu \sim \text{Gamma}(\nu/2, \nu/2)$, όπου η κατανομή γάμμα έχει συνάρτηση πυκνότητας πιθανότητας:

$$p(x | a, \beta) = \beta^a x^{a-1} \frac{\exp\{-\beta x\}}{\Gamma(a)}, x > 0, a > 0, \beta > 0 \quad (3.8)$$

ενώ $\Gamma(\cdot)$ είναι η συνάρτηση Γάμμα, $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$.

Ολοκληρώνοντας ως προς τ την από κοινού συνάρτηση πυκνότητας πιθανότητας, παίρνουμε την περιθώρια κατανομή. Η συνάρτηση πυκνότητας πιθανότητας της Student's- t κατανομής δίνεται από τη σχέση:

$$P(Y = (y_1, y_2, \dots, y_p)^T \mid \mu, \Sigma, \nu) = \frac{\Gamma\left(\frac{\nu+p}{2}\right) |\Sigma|^{-1/2}}{(\pi\nu)^{p/2} \Gamma\left(\frac{\nu}{2}\right) [1 + (Y - \mu)^T \Sigma^{-1} (Y - \mu)]^{(\nu+p)/2}} \quad (3.9)$$

όπου η ποσότητα $(Y - \mu)^T \Sigma^{-1} (Y - \mu)$ είναι η απόσταση Mahalanobis. Οι παράμετροι της κατανομής είναι η μέση τιμή μ , ο πίνακας συμμεταβλητότητας Σ και οι βαθμοί ελευθερίας ν . Για την παράμετρο ν μπορεί να δειχθεί ότι καθώς $\nu \rightarrow \infty$, $\tau \rightarrow 1$ με πιθανότητα 1, οπότε η Y γίνεται οριακά κανονική, $Y \sim N(\mu, \Sigma)$.

Η έννοια της παραμέτρου ν σχετίζεται με την ποσότητα θορύβου που υπάρχει στα δεδομένα. Είναι δηλαδή ένα μέτρο της απόκλισης των τελευταίων από το μοντέλο που τα περιγράφει. Όσο πιο μικρή τιμή έχει η παράμετρος ν τόσο πιο πολύς θόρυβος υπάρχει στα δεδομένα. Εδώ με την έννοια του θορύβου εννοούμε την απόκλιση από το μοντέλο περιγραφής των δεδομένων [7].

3.5. Κατάτμηση Φυσικών Εικόνων χωρίς Επίβλεψη

Όπως έχουμε αναφέρει παραπάνω, το πρόβλημα της κατάτμησης μιας φυσικής εικόνας είναι ένα από τα βασικότερα προβλήματα που έχουν απασχολήσει την επιστημονική κοινότητα. Μια «καλή» κατάτμηση πρέπει να ομαδοποιεί τα εικονοστοιχεία σε περιοχές που τα χαρακτηριστικά τους (χρώμα, φωτεινότητα, υφή) να είναι ομογενή και τα όριά τους απλά και ακριβή. Παρόλα αυτά, η κατάτμηση εικόνας είναι ένα διφορούμενο πρόβλημα για δύο βασικούς λόγους: Καταρχήν, η κατάτμηση μιας εικόνας σε περιοχές δεν είναι μοναδική, με την έννοια ότι τα εικονοστοιχεία μιας εικόνας μπορούν να ομαδοποιηθούν σε πολλές και διαφορετικές περιοχές. Επιπλέον, διαφορετικές περιοχές ή υφές έχουν διαφορετικές πολυπλοκότητες, πράγμα το οποίο δυσκολεύει τον καθορισμό του σωστού πλήθους ομάδων. Για παράδειγμα, εάν χρησιμοποιήσουμε κανονικές κατανομές για να μοντελοποιήσουμε τα χαρακτηριστικά διαφορετικών υφών, η κατανομή που θα χρησιμοποιήσουμε για μια απλή υφή, έχει προφανώς μικρότερη διάσταση απ' τη κανονική για μια πιο πολύπλοκη υφή.



Το μέσο πλήθος bits που χρειάζονται για να κωδικοποιήσουμε ένα τυχαίο διάνυσμα $v \in R^D$ με μια πολυδιάστατη κανονική κατανομή ως προς μια παραμόρφωση ε^2 , δίνεται από τη συνάρτηση ρυθμού-παραμόρφωσης, η οποία προσεγγίζεται ως:

$$R(\varepsilon) = \frac{1}{2} \log_2 \det(I + \frac{D}{\varepsilon^2} \Sigma) \quad (3.10)$$

όπου I ο μοναδιαίος πίνακας και Σ ο πίνακας συμμεταβλητότητας.

Εάν έχουμε N ανεξάρτητα διανύσματα $V = (v_1, v_2, \dots, v_N) \in R^{D \times N}$ που προέρχονται από την ίδια κανονική κατανομή, για τα οποία $\mu = \frac{1}{N} \sum_{i=1}^N v_i$ το μέσο τους και $\bar{V} = V - \mu * \mathbf{1}_{1 \times N}$ ο πίνακας συμμεταβλητότητας, τότε $\bar{\Sigma} = \frac{1}{N} \bar{V} \bar{V}^T$ μια προσέγγιση του Σ και:

$$R(\varepsilon, V) = \frac{1}{2} \log_2 \det(I + \frac{D}{\varepsilon^2 N} \bar{V} \bar{V}^T) \quad (3.11)$$

μια προσέγγιση της συνάρτησης ρυθμού - παραμόρφωσης.

Επομένως, για να κωδικοποιήσουμε τα N διανύσματα του V απαιτούνται $N * R(\varepsilon, V)$ bits, ή αλλιώς $D * R(\varepsilon, V)$ bits, δηλαδή ο ρυθμός κωδικοποίησης στους D πρωτεύοντες άξονες με πίνακα συμμεταβλητότητας $\bar{\Sigma} = \frac{1}{N} \bar{V} \bar{V}^T$. Εφόσον τα δεδομένα δεν έχουν μηδενικό μέσο, χρειαζόμαστε $\frac{D}{2} \log_2(1 + \frac{\mu^T \mu}{\varepsilon^2})$ bits για να κωδικοποιήσουμε το διάνυσμα των μέσων. Έτσι οδηγούμαστε στον τύπο 3.12 που μας δίνει τα συνολικά bits που χρειαζόμαστε για την κωδικοποίηση του συνόλου V :

$$L(V, \varepsilon) = \frac{N+D}{2} \log_2 \det(I + \frac{D}{\varepsilon^2 N} \bar{V} \bar{V}^T) + \frac{D}{2} \log_2(1 + \frac{\mu^T \mu}{\varepsilon^2}) \quad (3.12)$$

Ο τύπος 3.12 δεν ισχύει για μικτές κατανομές. Σ' αυτή την περίπτωση, η κωδικοποίηση του V γίνεται αν δούμε το V ως ένωση κάποιων υποσυνόλων $V = W_1 \cup W_2 \cup \dots \cup W_K$. Αν καθένα από αυτά τα υποσύνολα μοντελοποιείται με μια



Κανονική κατανομή, τότε το συνολικό πλήθος bits που απαιτούνται για να κωδικοποιήσουμε το V δίνεται από τον τύπο:

$$L^S(W_1, \dots, W_K) = \sum_{i=1}^K \{L(W_i) - |W_i| \log_2(|W_i|/N)\} \quad (3.13)$$

Η επιλογή της παραμόρφωσης ε γίνεται ελαχιστοποιώντας την παρακάτω συνάρτηση κόστους:

$$\varepsilon^* = \min_{\varepsilon \in E} \{L^S(V, \varepsilon) + \lambda ND \log_2(\varepsilon)\} \quad (3.14)$$

όπου λ μια παράμετρος που δίνει βάρη σε καθέναν από τους όρους $L^S(V, \varepsilon)$ και $ND \log_2(\varepsilon)$. Ο πρώτος όρος του παραπάνω αθροίσματος μειώνεται όσο το ε αυξάνεται, σε αντίθεση με την ποσότητα $ND \log_2(\varepsilon)$. Η παραπάνω συνάρτηση ψάχνει να βρει μια ισορροπία μεταξύ των bits που χρειάζονται για την κωδικοποίηση των δεδομένων και της παραμόρφωσης. Δεν θέλουμε πολλά bits, αλλά ούτε και μεγάλη παραμόρφωση των δεδομένων μας.

Η διαδικασία επιλογής της παραμόρφωσης ε , ξεκινάει από κάποιο μικρό ε και στη συνέχεια αυξάνει την τιμή του, επιλέγοντας από μια λίστα υποψήφιας τιμών, μέχρι η ελάχιστη *Kullback-Leibler* απόσταση μεταξύ των κατανομών των δύο υπερ-εικονοστοιχείων να είναι μεγαλύτερη από ένα προεπιλεγμένο κατώφλι γ : $\varepsilon^* = \min\{\varepsilon : d(\varepsilon) \geq \gamma\}$.

Η προέλευση της απόκλισης *Kullback-Leibler* (*KL divergence*) είναι από τη Θεωρία Πληροφορίας [2]. Για κατανομές $p(x)$, $q(x)$, η απόσταση KL ορίζεται ως:

$$d_{KL}(p \parallel q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \quad (3.15)$$

Αυξάνοντας το ε , μειώνουμε το πλήθος των υπερ-εικονοστοιχείων και το πλήθος των bits που χρειάζονται για την κωδικοποίηση. Η τελική κατάτμηση είναι αυτή που δίνει την πιο καθορισμένη τμηματοποίηση και ικανοποιεί την παραπάνω συνθήκη.



Για την κατανομή Student's- t , όπου μ και Σ , έχουμε τα αντίστοιχα διανύσματα των μέσων και τον πίνακα συμμεταβλητότητας που προκύπτουν από τη μοντελοποίηση των διανυσμάτων με την κατανομή αυτή.

Σε αυτό το σημείο θα περιγράψουμε τον αλγόριθμο με τον οποίο κάνουμε κατάτμηση μιας φυσικής εικόνας με τη βοήθεια της μεθόδου συμπίεσης με απώλειες [13]. Αρχικά, έχουμε σαν είσοδο μια εικόνα της οποίας τα εικονοστοιχεία θέλουμε να ομαδοποιήσουμε. Για την αναπαράσταση της RGB εικόνας, χρησιμοποιούμε τις Lab συντεταγμένες της. Η επιλογή αυτή γίνεται διότι μας διευκολύνει στην αναγνώριση της υψής, με τη βοήθεια κανονικών κατανομών.

Η εικόνα αυτή, χωρίζεται σε υπερ-εικονοστοιχεία, με την εφαρμογή του αλγορίθμου $N-cut$ και με τη χρήση των πιθανοτικών ορίων περιοχών. Η ομαδοποίηση των εικονοστοιχείων σε υπερ-εικονοστοιχεία, βοηθάει στην κατάτμηση για δύο βασικούς λόγους. Πρώτον, διότι μειώνεται το υπολογιστικό κόστος που απαιτείται για την κατάτμηση, εφόσον η επεξεργασία θα πρέπει να γίνει σε κάποιες εκατοντάδες εικονοστοιχείων και όχι στο σύνολο όλων των εικονοστοιχείων της εικόνας, και δεύτερον, γιατί χωρίς τη χρήση των υπερ-εικονοστοιχείων θα υπήρχε το πρόβλημα της ομαδοποίησης των εικονοστοιχείων που βρίσκονται κοντά στα όρια περιοχών, σε μία ξεχωριστή ομάδα.

Το επόμενο βήμα ύστερα από το χωρισμό της εικόνας σε υπερ-εικονοστοιχεία, είναι η κατασκευή μιας λίστας γειτνίασης των υπερ-εικονοστοιχείων αυτών. Δηλαδή, για κάθε υπερ-εικονοστοιχείο, αντιστοιχεί μια λίστα με όλα τα υπερ-εικονοστοιχεία με τα οποία συνορεύει. Αυτό θα μας βοηθήσει παρακάτω, που θα γίνει συνένωση του υπερ-εικονοστοιχείου με κάποιο/κάποια γειτονικά του. Η λίστα γειτνίασης υπερ-εικονοστοιχείων στο [13] αναφέρεται ως Region Adjacency Graph (RAG).

Εδώ θα πρέπει να αναφέρουμε ότι για να μοντελοποιήσουμε τα pixels του κάθε υπερ-εικονοστοιχείου χρησιμοποιούμε κανονικές κατανομές.



Στη συνέχεια, αντί να εφαρμόσουμε κάποια φίλτρα στην εικόνα μας, χρησιμοποιούμε τα *MRF*-χαρακτηριστικά, δημιουργώντας ένα παράθυρο αποκοπής γύρω από κάθε εικονοστοιχείο, μειώνοντας με αυτό τον τρόπο τη διάσταση των διανυσμάτων που περιγράφουν τα χαρακτηριστικά χρώματος και υφής για κάθε εικονοστοιχείο.

Το κριτήριο με βάση το οποίο θα συνενώσουμε ή όχι δύο γειτονικά υπερ-εικονοστοιχεία είναι τα bits τα οποία χρειάζονται για την κωδικοποίησή τους. Εάν τα bits που απαιτούνται πριν τη συνένωση είναι περισσότερα από τα bits που χρειάζονται αν γίνει η συνένωση, ως προς κάποια παραμόρφωση τότε η συνένωση των δύο γειτονικών υπερ-εικονοστοιχείων πρέπει να γίνει, αλλιώς όχι. Εάν γίνει η συνένωση, τότε τα συνολικά δεδομένα μας μοντελοποιούνται εκ νέου με τη βοήθεια μιας κανονικής κατανομής. Η διαδικασία αυτή επαναλαμβάνεται επαναληπτικά, όσο το ελάχιστο της απόστασης *Kullback-Leibler* δεν ξεπερνάει κάποιο κατώφλι εξ' αρχής ορισμένο. Στόχος μας είναι η κατάτμηση της εικόνας με τη χρήση όσο το δυνατόν λιγότερων bits.

Ο αλγόριθμος που περιγράφηκε παραπάνω, φαίνεται στο Σχήμα 3.5 και παράγει ως έξοδο τις ετικέτες των τελικών κατατμήσεων που προέκυψαν.

Επειδή η κανονική κατανομή δε μπορεί να μοντελοποιήσει τα εικονοστοιχεία τα οποία βρίσκονται κοντά στα σύνορα των υπερ-εικονοστοιχείων καθώς τους δίνει μηδενική πιθανότητα, αλλάζει το εύρος της (σ) για να τα μοντελοποιήσει και αυτά. Αυτό έχει ως αποτέλεσμα η καμπάνα της κανονικής κατανομής να αλλάζει, δίνοντας έτσι μικρότερες πιθανότητες στα εικονοστοιχεία που βρίσκονται στο εσωτερικό του υπερ-εικονοστοιχείου.

Μια λύση στο πρόβλημα αυτό δίνεται στο [8], στο οποίο δεν μοντελοποιούνται τα εικονοστοιχεία κοντά στα σύνορα, πράγμα το οποίο σημαίνει ότι δίνονται μηδενικές πιθανότητες στα εικονοστοιχεία που βρίσκονται κοντά στα σύνορα, αλλά το εύρος – επομένως και οι πιθανότητες που ανατίθενται στα εσωτερικά εικονοστοιχεία δεν αλλάζουν.



Μία άλλη λύση στο παραπάνω πρόβλημα, θα μπορούσε να είναι η χρήση της εύρωστης κατανομής Student's-*t*. Το βασικό πλεονέκτημα αυτής της κατανομής είναι ότι μπορεί να μοντελοποιήσει όλα τα στοιχεία που βρίσκονται κάτω από την καμπάνα της, ακόμα και αυτά που θεωρούνται outliers – στο δικό μας πρόβλημα, outliers θεωρούνται τα εικονοστοιχεία τα οποία βρίσκονται κοντά στα σύνορα των υπερ-εικονοστοιχείων. Επίσης, η κατανομή Student's-*t* δίνει μεγαλύτερη πιθανότητα στα εικονοστοιχεία που βρίσκονται κοντά στα συνορα να ανήκουν στο υπερ-εικονοστοιχείο αυτό.

Στον αλγόριθμο σχήματος 3.5 για κάθε υπερ-εικονοστοιχείο που δημιουργήθηκε μοντελοποιούμε τα στοιχεία του με την εύρωστη κατανομή Student's-*t*. Επίσης, στο στάδιο της συνένωσης δύο γειτονικών υπερ-εικονοστοιχείων, χρειάζεται να μοντελοποιήσουμε ξανά τα στοιχεία με την κατανομή Student's-*t*.

Αλλάζοντας την τιμή της παραμόρφωσης στους αλγορίθμους του σχήματος 3.5 παίρνουμε κάποια ιεραρχία κατατμήσεων και μπορούμε να δούμε το κατά πόσο καλά μπορεί να ανιχνευθεί η υφή. Με αυτό τον τρόπο επιλέγουμε την κατάλληλη τιμή για την παραμόρφωση η οποία ανιχνεύει καλύτερα την υφή της εικόνας μας. Όσο μικρότερη τιμή δεχθούμε να έχει η παραμόρφωση, τόσο πιο «καλή» αναμένουμε να είναι η κατάτμηση. Θα πρέπει επίσης να λάβουμε υπόψιν μας, ότι η τιμή της παραμόρφωσης θα πρέπει να είναι τέτοια ώστε να μπορούμε να διακρίνουμε περιοχές με διαφορετικές υφές, αλλά να μην είναι και πολύ μεγάλη, γιατί υπάρχει ο κίνδυνος να «χαλάσουμε» την κατάτμηση.

Μετά την εφαρμογή του αλγορίθμου αυτού, οδηγούμαστε στη βέλτιστη κατάτμηση, η οποία ελαχιστοποιεί το πλήθος των bits που είναι απαραίτητα για να κωδικοποιήσουμε τα δεδομένα, ως προς μια δοθείσα παραμόρφωση.



Είσοδοι: Εικόνα I σε Lab συντεταγμένες

Παράθυρο αποκοπής μεγέθους w

Φάσμα παραμορφώσεων E

Κατώφλι γ

1. Χωρισμός της εικόνας I σε υπερ-εικονοστοιχεία
2. Κατασκευή λίστας γειτνίασης, μία για κάθε υπερ-εικονοστοιχείο
3. Εφαρμογή παραθύρων αποκοπής μεγέθους w x w γύρω από κάθε εικονοστοιχείο και δημιουργία διανυσμάτων χαρακτηριστικών
4. Εφαρμογή της «Ανάλυσης σε πρωτεύουσες συνιστώσες» για τη μείωση της διάστασης των διανυσμάτων που προέκυψαν
5. Για όλες τις τιμές του φάσματος παραμορφώσεων E
 - a. Για κάθε υπερ-εικονοστοιχείο i
 1. Υπολόγισε τα συνολικά bits που απαιτούνται
 1. Για κάθε υπερ-εικονοστοιχείο j (γειτονικό του i)

Υπολόγισε τη διαφορά U_j των bits μετά τη συνένωση από τα bits (των δύο υπερ-εικονοστοιχείων) πριν τη συνένωση

Τέλος για
 2. Όσο το $\max\{U_j\} > 0$

Συνένωσε τα υπερεικονοστοιχεία i και j

Τέλος όσο
 - c. Αν $\min(\text{Kul_Leibl}(S_i, S_j, \epsilon)) \geq \gamma$

Σταμάτα

Τέλος αν

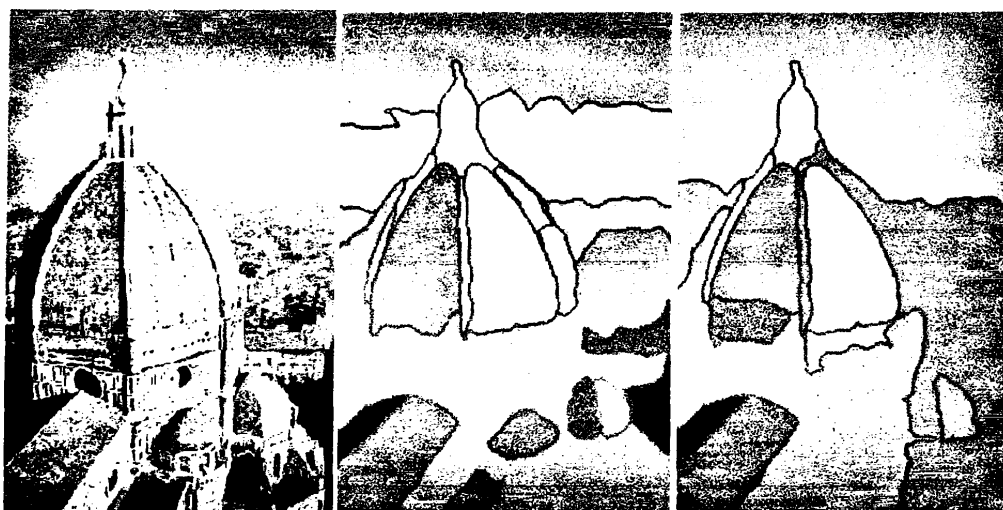
Έξοδος: Οι ετικέτες των ομάδων που προέκυψαν από την κατάτμηση

Σχήμα 3.5: Αλγόριθμος κατάτμησης με συμπίεση χωρίς απώλειες με την Κανονική κατανομή



3.6. Πειραματικά Αποτελέσματα

Στην υποενότητα αυτή, παραθέτουμε τα πειραματικά αποτελέσματα ύστερα από την εφαρμογή των παραπάνω αλγορίθμων, χρησιμοποιώντας Κανονική και Student's-t κατανομή σε καθένα από τα υπερεικονοστοιχεία της αρχικής εικόνας και σε όποια υπερ-εικονοστοιχεία συνενώνονται στην πορεία.

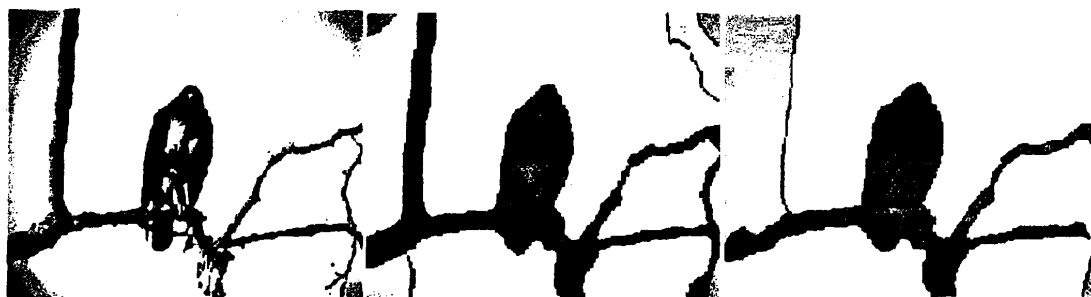


α)

β)

γ)

Σχήμα 3.6: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή

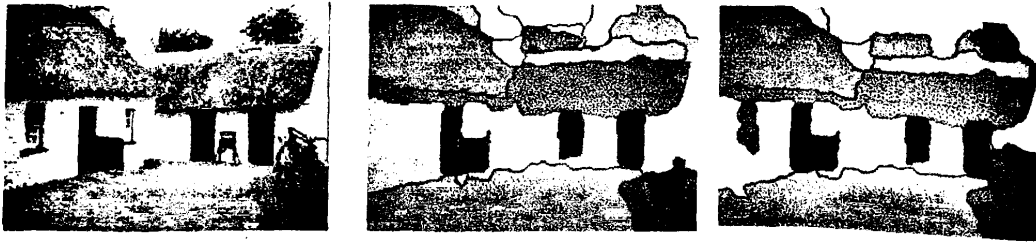


α)

β)

γ)

Σχήμα 3.7: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή



α)

β)

γ)

Σχήμα 3.8: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή.

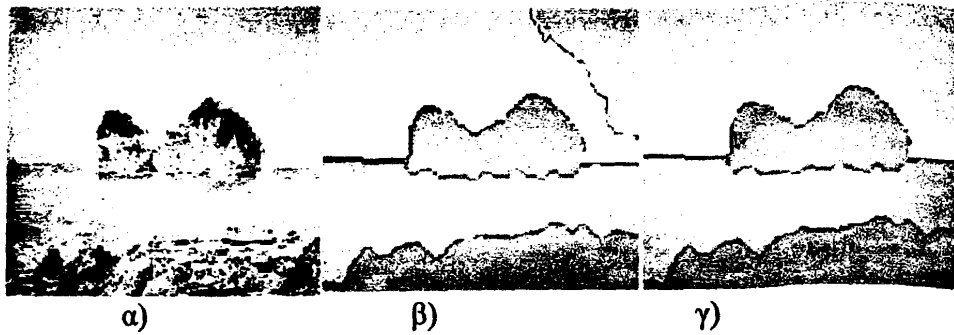


α)

β)

γ)

Σχήμα 3.9: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή



Σχήμα 3.10: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή

Τα παραπάνω πειραματικά αποτελέσματα προέκυψαν για κατώφλι $\gamma = 0.2$ και λίστα παραμορφώσεων:

$$\varepsilon \in \{0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.22, 0.24, 0.26, 0.28, 0.3, 0.35, 0.4, 0.45, 0.5\}$$

Παρακάτω παραθέτουμε τους δείκτες αξιολόγησης που προέκυψαν από τα πειράματά μας.

Στον πίνακα 3.1 φαίνονται οι δείκτες που προκύπτουν ύστερα από την εφαρμογή του παραπάνω αλγορίθμου στις εικόνες της βάσης του Berkeley, με τη χρήση της Κανονικής κατανομής.

Πίνακας 3.1: Δείκτες αξιολόγησης με τη χρήση της Κανονικής κατανομής

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.785	0.069	0.827
VoI	2.009	0.253	2.057
BDE	9.026	4.556	6.816
GCE	0.206	0.029	0.199

Στον πίνακα 3.2 φαίνονται οι δείκτες που προκύπτουν ύστερα από την εφαρμογή του παραπάνω αλγορίθμου στις εικόνες της βάσης του Berkeley, με τη χρήση της κατανομής Student's-t.

Πίνακας 3.2: Δείκτες αξιολόγησης με τη χρήση της κατανομής Student's-t

Δείκτης	Μέση τιμή	Τυπική Απόκλιση	Ενδιάμεσο
RI	0.600	0.055	0.622
VoI	3.081	0.298	3.022
BDE	20.862	5.388	17.890
GCE	0.355	0.052	0.368

Για τιμή κατωφλίου $\gamma = 0.1$ και για παραμόρφωση

$\varepsilon \in \{0.005, 0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.22, 0.24, 0.26, 0.28, 0.3, 0.35, 0.4, 0.45, 0.5\}$

Τα πειραματικά αποτελέσματα που προέκυψαν φαίνονται στα σχήματα 3.11, 3.12, 3.13, 3.14, 3.15, 3.17, 3.18, 3.19 και 3.20.

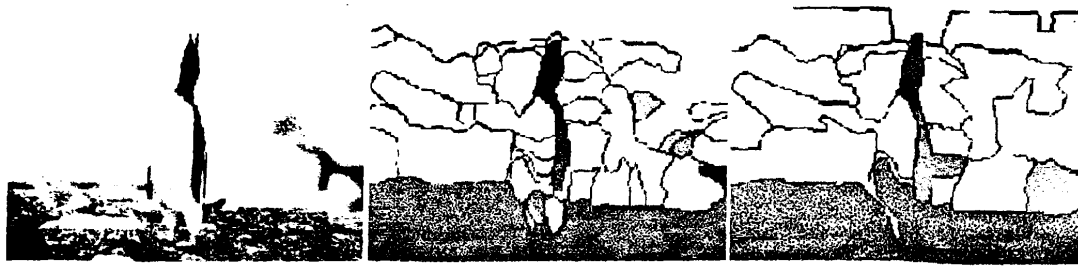


α)

β)

γ)

Σχήμα 3.11: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή



α)

β)

γ)

Σχήμα 3.12: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή



α)

β)

γ)

Σχήμα 3.13: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή

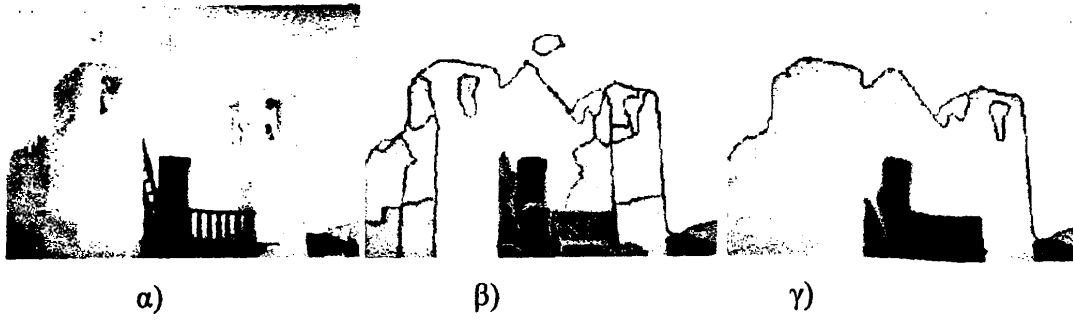


α)

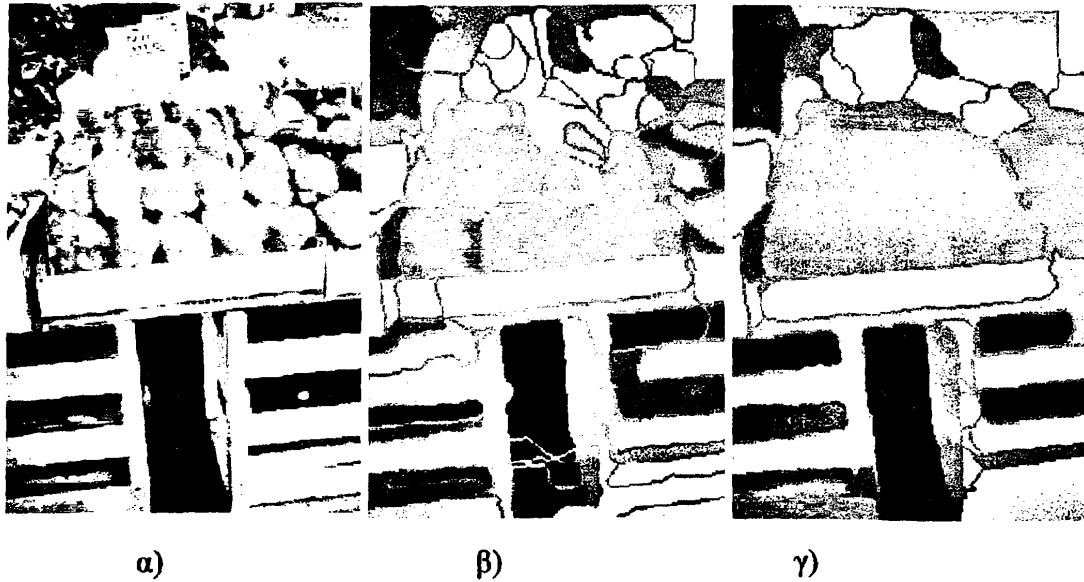
β)

γ)

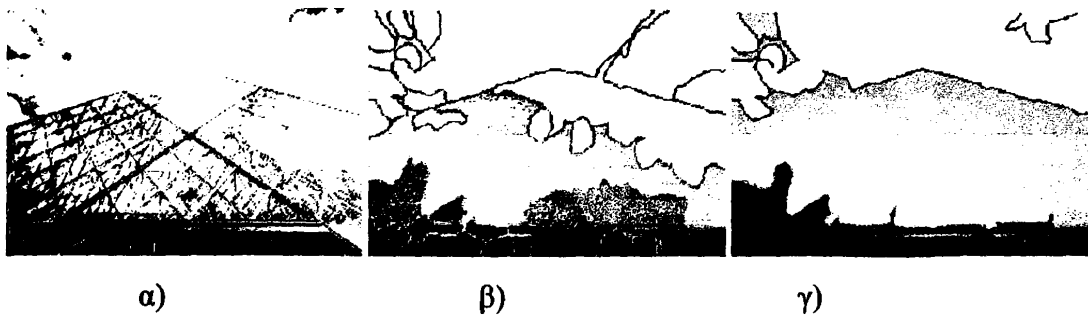
Σχήμα 3.14: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή



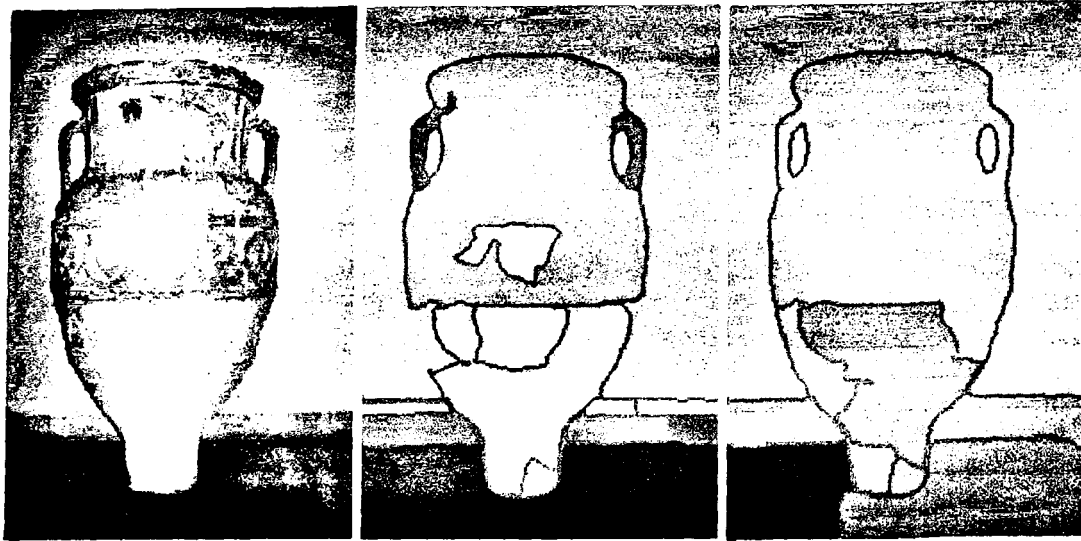
Σχήμα 3.15: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή



Σχήμα 3.16: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή



Σχήμα 3.17: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή



α)

β)

γ)

Σχήμα 3.18: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή

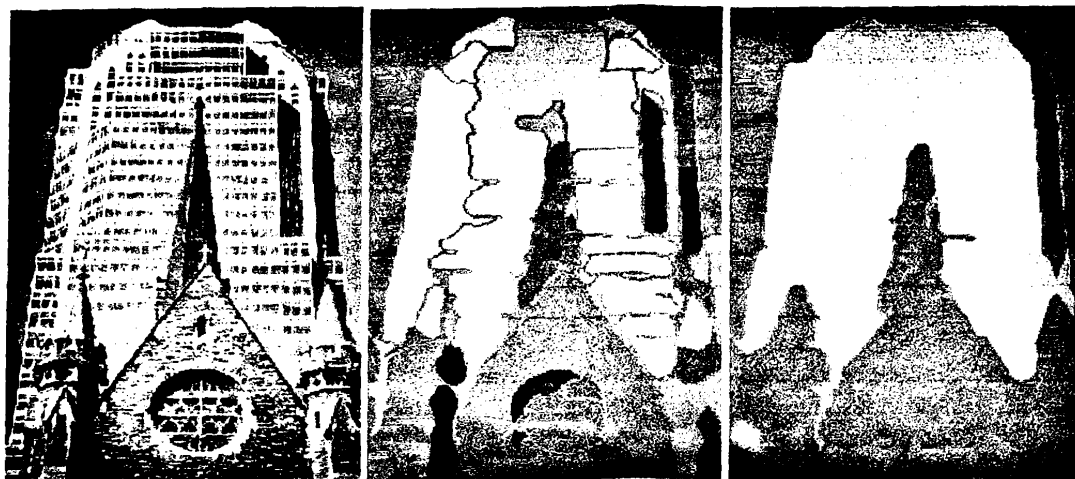


α)

β)

γ)

Σχήμα 3.19: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή



α)

β)

γ)

Σχήμα 3.20: α) Αρχική εικόνα, β) Κατάτμηση που προέκυψε από την Κανονική κατανομή, γ) Κατάτμηση που προέκυψε από τη Student's-t κατανομή

Στον πίνακα 3.3 φαίνονται οι δείκτες που προκύπτουν ύστερα από την εφαρμογή του παραπάνω αλγορίθμου σε 40 εικόνες της βάσης του Berkeley, με τη χρήση της Κανονικής κατανομής.

Πίνακας 3.3: Δείκτες αξιολόγησης με τη χρήση της Κανονικής κατανομής

Δείκτης	Μέση τιμή	Ενδιάμεσο
RI	0.707	0.745
VoI	4.142	4.279
BDE	12.859	12.426
GCE	0.339	0.341

Στον πίνακα 3.4 φαίνονται οι δείκτες που προκύπτουν ύστερα από την εφαρμογή του παραπάνω αλγορίθμου σε 40 εικόνες της βάσης του Berkeley, με τη χρήση της κατανομής Student's-t.

Πίνακας 3.4: Δείκτες αξιολόγησης με τη χρήση της κατανομής Student's-t

Δείκτης	Μέση τιμή	Ενδιάμεσο
RI	0.691	0.708
VoI	3.228	3.244
BDE	15.877	15.132
GCE	0.408	0.418

Από τα παραπάνω πειραματικά αποτελέσματα του πίνακα 3.4, παρατηρούμε ότι τα αποτελέσματα της κατάτμησης αν μοντελοποιήσουμε τα δεδομένα μας με τη χρήση της Student's-t κατανομής είναι αρκετά καλύτερα σε σύγκριση με αυτά του πίνακα 3.2.

Παρατηρούμε ότι η τιμή του *Rand Index* είναι σαφώς καλύτερη για κατώφλι 0.1, σε σύγκριση με την τιμή του για κατώφλι 0.2. Υπενθυμίζουμε, ότι όσο μεγαλύτερη τιμή έχει ο δείκτης *Rand Index*, τόσο καλύτερη κατάτμηση έχει επιτευχθεί.

Ο δείκτης αξιολόγησης *Variation of Information*, έχει περίπου την ίδια τιμή για κατώφλι 0.1 και 0.2, ενώ ο δείκτης *Boundary Detection Error* έχει μικρότερη, δηλαδή καλύτερη τιμή για κατώφλι 0.2.

Όσον αφορά τα αποτελέσματα που προέκυψαν ύστερα από τη μοντελοποίηση των δεδομένων με τη χρήση της Student's-t κατανομής σε σύγκριση με αυτά που προέκυψαν με τη χρήση της Κανονικής κατανομής, για τιμή κατωφλίου 0.2, έχουμε να κάνουμε τις παρακάτω παρατηρήσεις:

Η τιμή του δείκτη αξιολόγησης *Rand Index* είναι ελάχιστα καλύτερη με τη χρήση της Κανονικής κατανομής. Επίσης, ο δείκτης αξιολόγησης *Variation of Information* έχει καλύτερη τιμή με τη χρήση της Student's-t κατανομής.



Η Student's-*t* κατανομή, όπως αναφέραμε και παραπάνω, λόγω της μορφής της, θα αναμέναμε να δίνει καλύτερες κατατάξεις σε σύγκριση με αυτές που παράγονται με τη χρήση της Κανονικής κατανομής, καθώς δίνει μεγαλύτερες πιθανότητες στα εικονοστοιχεία τα οποία βρίσκονται κοντά στα σύνορα – και θεωρούνται outliers - να ανήκουν στη συγκεκριμένη ομάδα. Επίσης, η Student's-*t* κατανομή έχει την ιδιότητα να μην αλλάζει το εύρος της όπως συμβαίνει με την Κανονική κατανομή, αλλά να κρατάει το εύρος της σταθερό και να «σηκώνει» τα άκρα της, μοντελοποιώντας έτσι και τα δεδομένα τα οποία αποκλίνουν, δηλαδή τα εικονοστοιχεία που βρίσκονται κοντά στα σύνορα δύο ομάδων.

Ύστερα από τα παραπάνω πειραματικά αποτελέσματα, παρατηρούμε ότι η Student's-*t* κατανομή δίνει περίπου τις ίδιες τιμές στους δείκτες αξιολόγησης σε σύγκριση με την Κανονική κατανομή. Αν παρατηρήσουμε τα σχήματα 3.11 – 3.20, βλέπουμε ότι με μοντελοποίηση των εικονοστοιχείων με την Κανονική Κατανομή, δημιουργούνται πολλές περισσότερες ομάδες σε σύγκριση με τη χρήση της εύρωστης Student's-*t* κατανομής.

Όπως αναφέρθηκε και στην αρχή της εργασίας, δεν υπάρχει αντικειμενικός τρόπος ο οποίος θα μπορούσε να χαρακτηρίσει μια κατάτμηση «καλή» ή «κακή», ούτε και μπορούμε να βασιστούμε πλήρως στα αποτελέσματα τα οποία προέκυψαν από τους δείκτες αξιολόγησης. Επομένως, συγκρίνοντας τα οπτικά αποτελέσματα που προκύπτουν ύστερα από τη μοντελοποίηση των δεδομένων με τις δύο αυτές κατανομές, δε μπορούμε να πούμε με ασφάλεια ποια κατάτμηση είναι καλύτερη ή χειρότερη. Αυτό που παρατηρούμε όμως, είναι ότι η Student's-*t* κατανομή μας δίνει οπτικά καλύτερες ομάδες, διότι κατατάσσει στην ίδια ομάδα αντικείμενα που έχουν παρόμοια χαρακτηριστικά μεταξύ τους. Για παράδειγμα, αν παρατηρήσουμε το Σχήμα 3.16, βλέπουμε ότι η Student's-*t* κατανομή ομαδοποιεί όλα τα ομοειδή αντικείμενα σε μία μόνο ομάδα, ενώ η Κανονική κατανομή, ξεχωρίζει ένα – ένα τα αντικείμενα αυτά, πράγμα το οποίο μπορούμε να πούμε ότι δε μας δίνει κάποια σημαντική πληροφορία για την κατάτμησης μας. Το ίδιο συμβαίνει αν παρατηρήσουμε και τα υπόλοιπα σχήματα που προέκυψαν από τα πειραματικά μας αποτελέσματα.



ΚΕΦΑΛΑΙΟ 4. ΕΠΙΛΟΓΟΣ

Στην παρούσα μεταπτυχιακή εργασία, ασχοληθήκαμε με ένα βασικό τομέα του επιστημονικού πεδίου της επεξεργασίας εικόνας, αυτόν της κατάτμησης μιας εικόνας σε περιοχές. Παρουσιάστηκε μια μέθοδος κατάτμησης με ομαδοποίηση υπερεικονοστοιχείων, τα οποία είναι ομάδες από γειτονικά εικονοστοιχεία με συγγενή χαρακτηριστικά ως προς το χρώμα και την υφή και παρέχουν μια υπερκατάτμηση της εικόνας. Η αναπαράσταση των δεδομένων έγινε με πολυδιάστατα διανύσματα και με τη βοήθεια της ανάλυσης σε πρωτεύουσες συνιστώσες τα διανύσματα αυτά προβλήθηκαν σε χώρο μικρότερης διάστασης. Για κάθε υπερεικονοστοιχείο, τα νέα διανύσματα μοντελοποιήθηκαν από μία συνάρτηση πυκνότητας πιθανότητας και η κατάτμηση της εικόνας έγινε με βάση τη θεωρία ρυθμού-παραμόρφωσης, συγχωνεύοντας επαναληπτικά εκείνα τα υπερεικονοστοιχεία για τα οποία ο ρυθμός μετάδοσης της πληροφορίας που περιέχει η ένωσή τους είναι μέγιστος για δεδομένη παραμόρφωση. Σε αυτό το πλαίσιο, εξετάστηκε η μοντελοποίηση κάθε υπερεικονοστοιχείου από μία Κανονική κατανομή αλλά και από την εύρωστη κατανομή Student's-*t*. Αξιολογήσαμε τη μέθοδο αυτή, με τη χρήση δεικτών που συγκρίνουν τα αποτελέσματα των εξεταζόμενων τεχνικών με καταταμήσεις εθελοντών, αλλά και με βάση τα οπτικά αποτελέσματα τα οποία προέκυψαν από τα πειραματικά μας αποτελέσματα. Οι τιμές των δεικτών αξιολόγησης που προέκυψαν, ήταν περίπου ίδιες και στις δύο περιπτώσεις, πράγμα το οποίο δε μας βοήθησε να καταλήξουμε με ασφάλεια στο συμπέρασμα ότι κάποια από τις δύο καταταμήσεις είναι καλύτερη από την άλλη για τη βάση εικόνων του Berkeley. Παρατηρώντας όμως τις καταταμήσεις που προέκυψαν ύστερα από τη μοντελοποίηση των δεδομένων μας με την εύρωστη Student's-*t* κατανομή, καταλήγουμε στο συμπέρασμα ότι οι καταταμήσεις αυτές είναι καλύτερες σε σύγκριση με τη χρήση της Κανονικής κατανομής, καθώς ομοειδή αντικείμενα ομαδοποιούνται στην ίδια περιοχή, ενώ με τη



χρήση της Κανονικής κατανομής αντικείμενα με ομοειδή χαρακτηριστικά μεταξύ τους ως προς το χρώμα και την υφή, κατατάσσονται σε διαφορετικές ομάδες με αποτέλεσμα την ύπαρξη πολλών τμημάτων στην τελική μας κατάτμηση.



ΑΝΑΦΟΡΕΣ

- [1] Freixenet J., Munoz X., Rada D., Marti J., and Cuff X., "Yet another survey on image segmentation: Region and boundary information integration", University of Girona, Institute of Informatics and Applications, ECCV 2002, LNCS 2352, pp. 408-422, 2002.
- [2] Kullback S. and Leibler R.A., "On information and sufficiency", *Annals of Mathematical Statistics* 22: 79-86, 1951.
- [3] Martin D. R. , Fowlkes Ch. C., Tal D., Malik J., "A database of human segmented natural images and its application to evaluating segmentation algorithms", University of California, Berkeley, CA 94720.
- [4] Martin D. R., Fowlkes Ch. C., and Malik J., "Learning to detect natural image boundaries using local brightness, color, and texture cues", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.26, no. 5, pp. 530-549, May 2004.
- [5] Meila M., "Comparing clusterings - an axiomatic view", *Proceedings of the 22 nd International Conference on Machine Learning*, Bonn, Germany, 2005.
- [6] Pantofaru C., Hebert M., "A comparison of image segmentation algorithms", *The Robotics Institute Carnegie Mellon University Pittsburgh, Pennsylvania 15213*, pp. 1-25.
- [7] Peel D. and McLachlan G.J., *Robust mixture modeling using the t-distribution. Statistics and Computing*, 10:339-348, 2000.
- [8] Rao Sh. R., Mobahi H., Yang Al. Y., Sastry S. Sh. and Yi Ma, "Natural Image Segmentation with Adaptive Texture and Boundary Encoding", in *Proceedings of the Asian Conference on Computer Vision*, September 2009.
- [9] Shi J., Malik J., "Normalized cuts and image segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, no.8, pp. 888-905, August 2000.
- [10] Togneri R., Desilva Ch. J. S., "Fundamentals of information theory and coding design", *Discrete Mathematics and its applications*, Series Editor Kenneth H. Rosen, Chapman & Hall/Crc, pp. 88-96, 2002.



[11] Varma M., Zisserman A., "Texture classification: Are filter banks necessary?", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), Madison, Wisconsin, USA, Vol.2, pp. 691-698, 2003.

[12] Weiss Y., "Segmentation using eigenvectors: an unifying view", International Conference on Computer Vision, Berkeley, CA USA, 1994.

[13] Yang A. Y., Wright J., Sastry S., Yi Ma, "Unsupervised segmentation of natural images via lossy data compression", Computer Vision and Image Understanding, Vol. 110, no. 2, pp. 212-225, 8 March 2007.



ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Η Κωνσταντίνα Λουκοπούλου γεννήθηκε το 1985 στην Πάτρα Αχαΐας. Το 2002 εισήχθη στο Τμήμα Πληροφορικής του Πανεπιστημίου Ιωαννίνων. Το 2007 αποφοίτησε και συνέχισε τις σπουδές της στο Πρόγραμμα Μεταπτυχιακών Σπουδών του ίδιου τμήματος.

