

Κατάτμηση Φυσικής Εικόνας με Μεθόδους Επιβλεπόμενης και μη Επιβλεπόμενης Μάθησης

Χρήστος Παππάς

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

— ◆ —

Ιωάννινα, Φεβρουάριος 2011



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

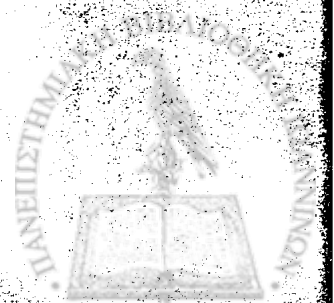
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF IOANNINA



ΒΙΒΛΙΟΘΗΚΗ
ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΙΩΑΝΝΙΝΩΝ



026000321469



Κατάτμηση φυσικής εικόνας με μεθόδους επιβλεπόμενης και
μη επιβλεπόμενης μάθησης

Η ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

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

από τον

Χρήστο Παππά

ως μέρος των Υποχρεώσεων για τη λήψη του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ
ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ
ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ – ΕΦΑΡΜΟΓΕΣ

Φεβρουάριος 2011



ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τον επιβλέποντα κ. Νίκου Χριστόφορο, Επίκουρο Καθηγητή του τμ. Πληροφορικής του Πανεπιστημίου Ιωαννίνων, για την πολύ καλή συνεργασία που είχαμε κατά την εκπόνηση της διπλωματικής μου εργασίας. Ουσιαστική ήταν και η βοήθεια του κ. Λύκα Αριστεΐδη, Αναπληρωτή Καθηγητή του τμ. Πληροφορικής του Πανεπιστημίου Ιωαννίνων, του οποίου οι συμβουλές ήταν πολύτιμες και τον οποίο ευχαριστώ βαθύτατα. Τέλος να ευχαριστήσω την οικογένεια μου για τη συνεχή υποστήριξη της καθόλη τη διάρκεια της πραγματοποίησης της εργασίας.



ΠΕΡΙΕΧΟΜΕΝΑ

1	Εισαγωγή	10
1.1.	Εισαγωγή	10
2	Θεωρητικό υπόβαθρο	12
2.1	Εισαγωγή	12
2.2	MRF χαρακτηριστικά	13
2.3	SIFT χαρακτηριστικά	13
2.4	Ανάλυση σε πρωτεύουσες συνιστώσες	15
2.4.1	Μαθηματικές λεπτομέρειες	15
2.4.2	Διαπιστώσεις	17
2.5	Φασματική ομαδοποίηση σε επίπεδο εικονοστοιχείου	18
2.5.1	Φασματική ομαδοποίηση με χρήση του Λαπλασιανού πίνακα	18
2.5.2	Φασματική ομαδοποίηση με χρήση του πίνακα ομοιότητας	23
2.5.3	Παραδείγματα υπερεικονοστοιχείων	25
2.6	Φασματική ομαδοποίηση σε επίπεδο υπερεικονοστοιχείου	25
2.7	Ταξινομητές	28
2.7.1	Απόκλιση Kullback–Leibler	28
2.7.2	Μηχανές διανυσμάτων υποστήριξης	29
3	Κατάτμηση εικόνας με μεθόδους μη επιβλεπόμενης μάθησης	35
3.1	Εισαγωγή	35
3.2	Βάση του Berkeley	35
3.3	Αλγόριθμος	36
3.4	Πειραματικά αποτελέσματα	37
4	Κατάτμηση εικόνας με μεθόδους επιβλεπόμενης μάθησης	55
4.1	Εισαγωγή	55
4.2	Σημασιολογική κατάτμηση	55
4.3	Δάση στοιχείων υψής με εννοιολογικά χαρακτηριστικά	56
4.3.1	Τυχαία δέντρα απόφασης	57
4.3.2	Εκπαίδευση STF	58
4.3.3	Κατηγοριοποίηση STF	59
4.4	Βάση MSRC21	60



4.5	Αλγόριθμος	60
4.6	Πειραματικά Αποτελέσματα	64
5	Επίλογος	71
5.1	Επίλογος	71



ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

2.1	MRF χαρακτηριστικά.	14
2.2	Οι κλίσεις της εικόνας γύρω από το εικονοστοιχείο και ο περιγραφέας του. Το σχήμα αναπαράχθηκε από το [8].	15
2.3	Εικόνες και υπερεικονοστοιχεία.	26
3.1	Εικόνες της βάσης του Berkeley και τα δοθέντα υπερεικονοστοιχεία τους.	37
3.2	Η κατάτμηση της εικόνας "Αστερίας" σε 3 ομάδες με χρήση της απόκλισης KL. Βέλτιστη παράμετρος $a=0.3$	41
3.3	Η κατάτμηση της εικόνας "Αστερίας" σε 3 ομάδες με χρήση της απόστασης κοντινότερου γείτονα. Βέλτιστη παράμετρος $a=0.7$	41
3.4	Η κατάτμηση της εικόνας "Άλογα" σε 3 ομάδες με χρήση της απόκλισης KL. Βέλτιστη παράμετρος $a=0.9$	42
3.5	Η κατάτμηση της εικόνας "Άλογα" σε 3 ομάδες με χρήση της απόστασης κοντινότερου γείτονα. Βέλτιστη παράμετρος $a=0.9$	42
3.6	Η κατάτμηση της εικόνας "Ελέφαντας" σε 5 ομάδες με χρήση της απόκλισης KL. Βέλτιστη παράμετρος $a=0.7$	43
3.7	Η κατάτμηση της εικόνας "Ελέφαντας" σε 5 ομάδες με χρήση της απόστασης κοντινότερου γείτονα. Βέλτιστη παράμετρος $a=0.3$	43
3.8	Η κατάτμηση της εικόνας "Αρκούδα" σε 5 ομάδες με χρήση της απόκλισης KL. Βέλτιστη παράμετρος $a=0.7$	44
3.9	Η κατάτμηση της εικόνας "Αρκούδα" σε 5 ομάδες με χρήση της απόστασης κοντινότερου γείτονα. Βέλτιστη παράμετρος $a=0.5$	44
3.10	Η κατάτμηση της εικόνας "Αγρόκτημα" σε 8 ομάδες με χρήση της απόκλισης KL. Βέλτιστη παράμετρος $a=0.7$	45
3.11	Η κατάτμηση της εικόνας "Αγρόκτημα" σε 8 ομάδες με χρήση της απόστασης κοντινότερου γείτονα. Βέλτιστη παράμετρος $a=0.9$	45
3.12	Αξιολόγηση των κατατμήσεων με τη μέθοδο Ncut και απόκλιση KL. Ο δείκτης BDE. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος.	47



- 3.13 Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόκλιση KL. Ο δείκτης **PRI**. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος. 47
- 3.14 Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόκλιση KL. Ο δείκτης **VOI**. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος. 48
- 3.15 Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόκλιση KL. Ο δείκτης **GCE**. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος. 48
- 3.16 Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόσταση κοντινότερου γείτονα. Ο δείκτης **BDE**. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος. 50
- 3.17 Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόσταση κοντινότερου γείτονα. Ο δείκτης **PRI**. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος. 50
- 3.18 Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόσταση κοντινότερου γείτονα. Ο δείκτης **VOI**. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος. 51
- 3.19 Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόσταση κοντινότερου γείτονα. Ο δείκτης **GCE**. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος. 51
- 3.20 Αξιολόγηση των κατατιμήσεων με τη μέθοδο φασματικής ομαδοποίησης με χρήση του πίνακα ομοιότητας και απόκλιση KL. Ο δείκτης **BDE**. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος. 53
- 3.21 Αξιολόγηση των κατατιμήσεων με τη μέθοδο φασματικής ομαδοποίησης με χρήση του πίνακα ομοιότητας και απόκλιση KL. Ο δείκτης **PRI**. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος. 53



3.22	Αξιολόγηση των κατατιμήσεων με τη μέθοδο φασματικής ομαδοποίησης με χρήση του πίνακα ομοιότητας και απόκλιση KL. Ο δείκτης VOI. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος.	54
3.23	Αξιολόγηση των κατατιμήσεων με τη μέθοδο φασματικής ομαδοποίησης με χρήση του πίνακα ομοιότητας και απόκλιση KL. Ο δείκτης GCE. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος.	54
4.1	Αρχική εικόνα και η σημασιολογική της κατάτμηση. Το σχήμα αναπαράχθηκε από το [6].	56
4.2	Απλό σημασιολογικό δέντρο απόφασης. Το σχήμα αναπαράχθηκε από το [6]. . .	58
4.3	Εικόνες της βάσης MSRC21 και οι κατηγορίες των εικονοστοιχείων τους.	61
4.4	Σημασιολογική κατάτμηση 1.	65
4.5	Σημασιολογική κατάτμηση 2.	66
4.6	Σημασιολογική κατάτμηση 3.	67
4.7	Σημασιολογική κατάτμηση 4.	68
4.8	Σημασιολογική κατάτμηση 5.	69



ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

- 3.1 Αξιολόγηση των κατατμήσεων με τη μέθοδο Ncut και απόκλιση KL. Συμβολίζουμε με M τον αριθμό ομάδων και με a τη σταθερά γειτνίασης των υπερεικονοστοιχείων. 46
- 3.2 Αξιολόγηση των κατατμήσεων με τη μέθοδο Ncut και απόσταση κοντινότερου γείτονα. Συμβολίζουμε με M τον αριθμό ομάδων και με a τη σταθερά γειτνίασης των υπερεικονοστοιχείων. 49
- 3.3 Αξιολόγηση των κατατμήσεων με τη μέθοδο φασματικής ομαδοποίησης με χρήση του πίνακα ομοιότητας και απόκλιση KL. Συμβολίζουμε με M τον αριθμό ομάδων και με a τη σταθερά γειτνίασης των υπερεικονοστοιχείων. . . 52
- 4.1 Αξιολόγηση των σημασιολογικών κατατμήσεων με ταξινομητή 1-NN και απόκλιση KL. Συμβολίζουμε με 'supr' τα υπερεικονοστοιχεία και με 'pixels' τα εικονοστοιχεία. 70
- 4.2 Αξιολόγηση των σημασιολογικών κατατμήσεων με ταξινομητή SVM. Συμβολίζουμε με 'supr' τα υπερεικονοστοιχεία και με 'pixels' τα εικονοστοιχεία. 70



ΕΥΡΕΤΗΡΙΟ ΑΛΓΟΡΙΘΜΩΝ

1	Φασματική ομαδοποίηση σε επίπεδο εικονοστοιχείου με χρήση του Λαπλασιανού πίνακα (Ncut)	22
2	Φασματική ομαδοποίηση σε επίπεδο εικονοστοιχείου με χρήση του πίνακα ομοιότητας	25
3	Φασματική ομαδοποίηση σε επίπεδο υπερεικονοστοιχείου με χρήση του Λαπλασιανού πίνακα	27
4	Φασματική ομαδοποίηση σε επίπεδο υπερεικονοστοιχείου με χρήση του πίνακα ομοιότητας	27
5	Αλγόριθμος μη επιβλεπόμενης μάθησης	38
6	Αλγόριθμος επιβλεπόμενης μάθησης με ταξινομητή KL	63
7	Αλγόριθμος επιβλεπόμενης μάθησης με ταξινομητή SVM	63



ΠΕΡΙΛΗΨΗ

Χρήστος Παππάς του Δημητρίου και της Κωνσταντίας. MSc, Τμήμα Πληροφορικής Πανεπιστημίου Ιωαννίνων, Φεβρουάριος, 2011. Κατάτμηση φυσικής εικόνας με μεθόδους επιβλεπόμενης και μη επιβλεπόμενης μάθησης. Επιβλέποντας: Χριστόφορος Νίχου

Η παρούσα εργασία πραγματεύεται το αντικείμενο της κατάτμησης φυσικής εικόνας. Στόχος της κατάτμησης εικόνας είναι να απλοποιήσει ή να αλλάξει την αναπαράσταση μιας εικόνας διαχωρίζοντας τη σε ομοιογενείς περιοχές, που αναμένεται να αντιστοιχούν σε αντικείμενα και έννοιες του φυσικού κόσμου. Οι περιοχές που προκύπτουν είναι, συνεπώς, ευκολότερο να προσδιοριστούν ανεξάρτητα ως οντότητες στο πλαίσιο της εικόνας καθώς και να υποστούν περαιτέρω εσωτερική επεξεργασία ανάλογα με την εφαρμογή. Μερικές πρακτικές εφαρμογές της κατάτμησης εικόνας είναι οι ιατρικές απεικονίσεις, ο εντοπισμός αντικειμένων, ο αυτόματος έλεγχος συστημάτων κυκλοφορίας κ.λ.π. Στο πρώτο μέρος της εργασίας αναλύεται το θεωρητικό υπόβαθρο στο οποίο στηρίζονται οι μέθοδοι επιβλεπόμενης και μη επιβλεπόμενης μάθησης που χρησιμοποιούνται. Συγκεκριμένα, αρχικά μελετώνται χαρακτηριστικά που αναπαριστούν αποδοτικά φωτεινότητα, χρώμα και υφή και περιγράφουν ένα εικονοστοιχείο, χρησιμοποιώντας την πληροφορία της τοπικής γειτονιάς του. Έπειτα, παρουσιάζεται η ανάλυση πρωτευουσών συνιστωσών (PCA) που αποτελεί μία αξιόπιστη τεχνική για μείωση της διάστασης των δεδομένων του προβλήματος. Στη συνέχεια, περιγράφεται ο αλγόριθμος φασματικής ομαδοποίησης τόσο σε επίπεδο εικονοστοιχείου όσο και σε επίπεδο υπερεικονοστοιχείου. Τα υπερεικονοστοιχεία ορίζονται ως συμπαγείς περιοχές ομαδοποίησης των εικονοστοιχείων μιας εικόνας και αποτελούν τη βάση των μεθόδων μάθησης. Τέλος, παρουσιάζονται οι ταξινομητές που χρησιμοποιήθηκαν. Στο δεύτερο μέρος της εργασίας παρουσιάζονται οι αλγόριθμοι μη επιβλεπόμενης μάθησης ως μέθοδοι ομαδοποίησης των υπερεικονοστοιχείων μιας εικόνας σε ομοιογενείς περιοχές με τη χρήση των χαρακτηριστικών που μελετήθηκαν στο πρώτο μέρος της εργασίας. Ακολουθεί η εφαρμογή τους στη βάση εικόνων του Berkeley. Στο τρίτο μέρος της εργασίας αναλύονται οι αλγόριθμοι επιβλεπόμενης μάθησης ως μέθοδοι ανάθεσης κατηγοριών σε υπερεικονοστοιχεία μιας εικόνας με τη χρήση αντίστοιχων χαρακτηριστικών από το πρώτο μέρος. Επιδιώκεται η σημασιολογική κατάτμηση της εικόνας σε ευρύτερα ομοιογενείς περιοχές, οι οποίες προκύπτουν με την ενοποίηση ομοίως κατηγοριοποιημένων υπερεικονοστοιχείων. Ακολουθεί η εφαρμογή τους στη βάση σημασιολογικών εικόνων MSRC21. Τελικά, οι μέθοδοι επιβλεπόμενης και μη επιβλεπόμενης μάθησης που εφαρμόστηκαν προσέδωσαν ικανοποιητικά αποτελέσματα κατατμήσεων τόσο οπτικά, καθώς προκύπτουν υποπεριοχές της εικόνας που αντιστοιχούν σε οντότητες του φυσικού κόσμου, όσο και βάσει των δεικτών αξιολόγησής.



EXTENDED ABSTRACT IN ENGLISH

Pappas, Christos. MSc, Computer Science Department, University of Ioannina, Greece. February, 2011. Natural image segmentation with supervised and unsupervised learning methods. Thesis Supervisor: Christophoros Nikou.

The research topic of this dissertation addresses the problem of natural image segmentation. The objective of image segmentation is to simplify or change the representation of an image by separating it into homogeneous regions expected to represent objects and concepts of the natural world. The resulting regions are therefore easier to identify as independent entities in the image as well as undergo further internal processing depending on the application needs. Some practical applications of image segmentation are medical imaging, object detection, automated traffic control systems, etc. The first part of this dissertation deals with the theoretical background underlying the methods of supervised and unsupervised learning used. Primarily, we study features that effectively represent brightness, color and texture, and describe one pixel using the information of its local neighborhood. We also present principal component analysis (PCA) which is a reliable technique for reducing the dimension of the problem data. Then, spectral clustering algorithm is described in both pixel and superpixel level. Superpixels are defined as compact grouping regions of pixels and form the basis of the learning methods described. Finally, we analyze the classifiers used for training and testing. The second part of this dissertation presents the unsupervised learning algorithms. The aim of these methods is to cluster image superpixels into homogeneous regions by using the features of the first part of our work. Next follows their implementation on the Berkeley image segmentation dataset. The third part of this dissertation presents the supervised learning algorithms. The object of these methods is to assign a category to each image superpixel by using, respectively, the features of the first part. Semantic image segmentation in widely homogeneous areas is intended by joining similarly classified superpixels. Their implementation on the MSRC21 image database follows. Finally, the supervised and unsupervised learning methods applied result in satisfactory image segmentations based on both visual representation, as the resulting image subregions correspond to entities of the natural world, as well as segmentation evaluation.



ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

1.1 Εισαγωγή

1.1 Εισαγωγή

Η κατάτμηση εικόνας στην υπολογιστική όραση αναφέρεται στη διαδικασία του διαχωρισμού μιας ψηφιακής εικόνας σε πολλαπλές περιοχές. Ο σκοπός της κατάτμησης είναι να απλοποιήσουμε και/ή να αλλάξουμε την αναπαράσταση μιας εικόνας σε κάτι το οποίο είναι πιο κατανοητό και ευκολότερο να αναλυθεί περαιτέρω. Η κατάτμηση εικόνας χρησιμοποιείται συνήθως για να εντοπίσουμε αντικείμενα και όρια περιοχών σε εικόνες. Ειδικότερα, η κατάτμηση είναι η διαδικασία της ανάθεσης μιας "ετικέτας" σε κάθε εικονοστοιχείο σε μια εικόνα, έτσι ώστε εικονοστοιχεία με την ίδια ετικέτα μοιράζονται ορισμένα οπτικά χαρακτηριστικά.

Το αποτέλεσμα της κατάτμησης είναι ένα σύνολο από τμήματα που συλλογικά καλύπτουν ολόκληρη την εικόνα ή ένα σύνολο από περιγράμματα που εξάγονται από την εικόνα (για παράδειγμα ανίχνευση ακμής). Τα εικονοστοιχεία που ανήκουν σε μια περιοχή είναι παρόμοια βάσει κάποιων χαρακτηριστικών και ιδιοτήτων, όπως χρώμα, φωτεινότητα και υφή. Διαφορετικές περιοχές εικονοστοιχείων πρέπει να είναι διακριτές μεταξύ τους όσον αφορά τα αντίστοιχά τους χαρακτηριστικά.

Μερικές από τις πρακτικές εφαρμογές της κατάτμησης εικόνας είναι:

- Ιατρικές απεικονίσεις
 - Εντοπισμός όγκων και άλλων παθολογιών.
 - Καθοδηγούμενη ιατρική επέμβαση.
 - Διάγνωση.
 - Μελέτη της ανατομικής δομής.



- Εντοπισμός αντικειμένων σε δορυφορικές εικόνες (για παράδειγμα δρόμοι, δάση κ.λ.π.).
- Αναγνώριση προσώπου.
- Αναγνώριση δακτυλικών αποτυπωμάτων.
- Αυτόματος έλεγχος συστημάτων κυκλοφορίας.
- Έλεγχος βιομηχανικής διαδικασίας ή δραστηριότητας.

Όσον αφορά την ανίχνευση αντικειμένων, η κατάτμηση της εικόνας μπορεί να είναι δύσκολη για διάφορους λόγους:

- απώλεια πληροφορίας κατά την προβολή από τις τρεις διαστάσεις του πραγματικού κόσμου στις δύο διαστάσεις της εικόνας.
- θόρυβος κατά την αποτύπωση της εικόνας.
- μη συμπαγή αντικείμενα.
- ποικιλομορφία αντικειμένων της ίδιας κατηγορίας.
- αλλαγή στον φωτισμό της σκηνής.
- απαιτήσεις πραγματικού χρόνου.

Αρκετοί γενικής χρήσης αλγόριθμοι και τεχνικές έχουν αναπτυχθεί για την κατάτμηση εικόνας. Δεδομένου ότι δεν υπάρχει καμία γενική λύση στο πρόβλημα κατάτμησης εικόνας, αυτές οι τεχνικές πρέπει συχνά να συνδυαστούν με τη γνώση περιοχών προκειμένου να λυθεί αποτελεσματικά ένα πρόβλημα κατάτμησης εικόνας για μια περιοχή προβλήματος.

Στο κεφάλαιο 2 αναλύεται σε βάθος το θεωρητικό υπόβαθρο πάνω στο οποίο στηρίζονται οι μέθοδοι κατάτμησης που περιγράφονται στα κεφάλαια 3 και 4. Στο κεφάλαιο 3 παρουσιάζονται οι αλγόριθμοι μη εποπτευόμενης μάθησης και η εφαρμογή τους για την ανίχνευση ομοιογενών περιοχών σε εικόνες. Στο κεφάλαιο 4 περιγράφονται οι αλγόριθμοι εποπτευόμενης μάθησης και η εφαρμογή τους για εντοπισμό συγκεκριμένων κατηγοριών σε εικόνες.



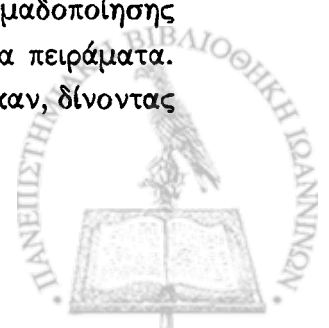
ΚΕΦΑΛΑΙΟ 2

ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

-
- 2.1 Εισαγωγή
 - 2.2 MRF χαρακτηριστικά
 - 2.3 SIFT χαρακτηριστικά
 - 2.4 Ανάλυση σε πρωτεύουσες συνιστώσες
 - 2.5 Φασματική ομαδοποίηση σε επίπεδο εικονοστοιχείου
 - 2.6 Φασματική ομαδοποίηση σε επίπεδο υπερεικονοστοιχείου
 - 2.7 Ταξινομητές
-

2.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται αναλυτικά το θεωρητικό υπόβαθρο που απαιτείται για την περιγραφή των μεθόδων κατάτμησης των κεφαλαίων 3 και 4. Στην παράγραφο 2.2 μελετώνται τα MRF χαρακτηριστικά που περιγράφουν ένα εικονοστοιχείο (pixel), χρησιμοποιώντας την πληροφορία των γειτονικών του εικονοστοιχείων, και τα οποία κατηγοριοποιούν υφή αρκετές φορές καλύτερα από την εφαρμογή αντίστοιχων φίλτρων ανίχνευσης υφής. Στην παράγραφο 2.3 περιγράφονται τα SIFT χαρακτηριστικά που αφορούν κάθε εικονοστοιχείο και εμπεριέχουν πληροφορία των γειτονικών του εικονοστοιχείων. Στην παράγραφο 2.4 παρουσιάζεται η ανάλυση πρωτευουσών συνιστωσών (PCA). Στην παράγραφο 2.5 περιγράφονται οι αλγόριθμοι φασματικής ομαδοποίησης σε επίπεδο εικονοστοιχείου, οι οποίοι χρησιμοποιούνται για την ομαδοποίηση των επιμέρους εικονοστοιχείων σε υπερεικονοστοιχεία [22], [9]. Στην παράγραφο 2.6 αναλύονται οι αλγόριθμοι της φασματικής ομαδοποίησης σε επίπεδο υπερεικονοστοιχείου (superpixel), οι οποίοι χρησιμοποιούνται στα πειράματα. Τέλος, στην παράγραφο 2.7 περιγράφονται οι ταξινομητές που χρησιμοποιήθηκαν, δίνοντας



έμφαση στην παραμετροποίηση και την εφαρμογή τους στο σύνολο πειραμάτων των μεθόδων των κεφαλαίων 3 και 4.

2.2 MRF χαρακτηριστικά

Τα MRF (Markov Random Field) χαρακτηριστικά έχει αποδειχτεί [23] ότι μπορεί να αποτελέσουν ένα πολύ σημαντικό εργαλείο για εντοπισμό και κατηγοριοποίηση υφής και μάλιστα μπορούν να ξεπεράσουν την απόδοση των συστοιχιών φίλτρων [5] σε μεγάλο βαθμό.

Η βασική ιδέα είναι ότι μπορούμε να περιγράψουμε ένα εικονοστοιχείο, αξιοποιώντας την πληρόφωρα της γειτονιάς του. Τότε, η υφή μπορεί να αναπαρασταθεί χρησιμοποιώντας την από κοινού κατανομή των τιμών φωτεινότητας πάνω σε εξαιρετικά συμπαγείς γειτονιές των εικονοστοιχείων και, επομένως, ομαδοποιούμε τα εικονοστοιχεία που χαρακτηρίζονται από παρόμοια υφή. Συνήθως χρησιμοποιούνται όλες οι τιμές των εικονοστοιχείων της τοπικής γειτονιάς του εικονοστοιχείου που μας ενδιαφέρει, όπου το μέγεθος της γειτονιάς προτείνεται να είναι ένα παράθυρο μεγέθους 3×3 έως 7×7 . Γενικότερα, θεωρώντας τετραγωνικό μέγεθος παραθύρου $N \times N$ γύρω από το σημείο, σχηματίζεται διάνυσμα τιμών στον $[N^2]$ -διάστατο χώρο χαρακτηριστικών για εικόνα με αποχρώσεις του γκρι και διάνυσμα τιμών στον $[3 * N^2]$ -διάστατο χώρο χαρακτηριστικών για έγχρωμη εικόνα. Αξιοποιούμε το διάνυσμα αυτό ως την πληροφορία που χαρακτηρίζει το εικονοστοιχείο.

Η παρακάτω εξίσωση συνοψίζει την έννοια των MRF χαρακτηριστικών

$$p(I(x_c)|I(x), \forall x \neq x_c) = p(I(x_c)|I(x), x \in \mathcal{N}(x_c))$$

όπου το x_c είναι ένα σημείο στον διδιάστατο πλέγμα ακεραίων στο οποίο έχει οριστεί η εικόνα I και $\mathcal{N}(x_c)$ είναι το σύνολο των σημείων της γειτονιάς του x_c . Στην περίπτωση μας, το \mathcal{N} έχει οριστεί να είναι το τετραγωνικό πλέγμα γειτονιάς μεγέθους $N \times N$, συμπεριλαμβανομένου και του κεντρικού εικονοστοιχείου στη θέση x_c .

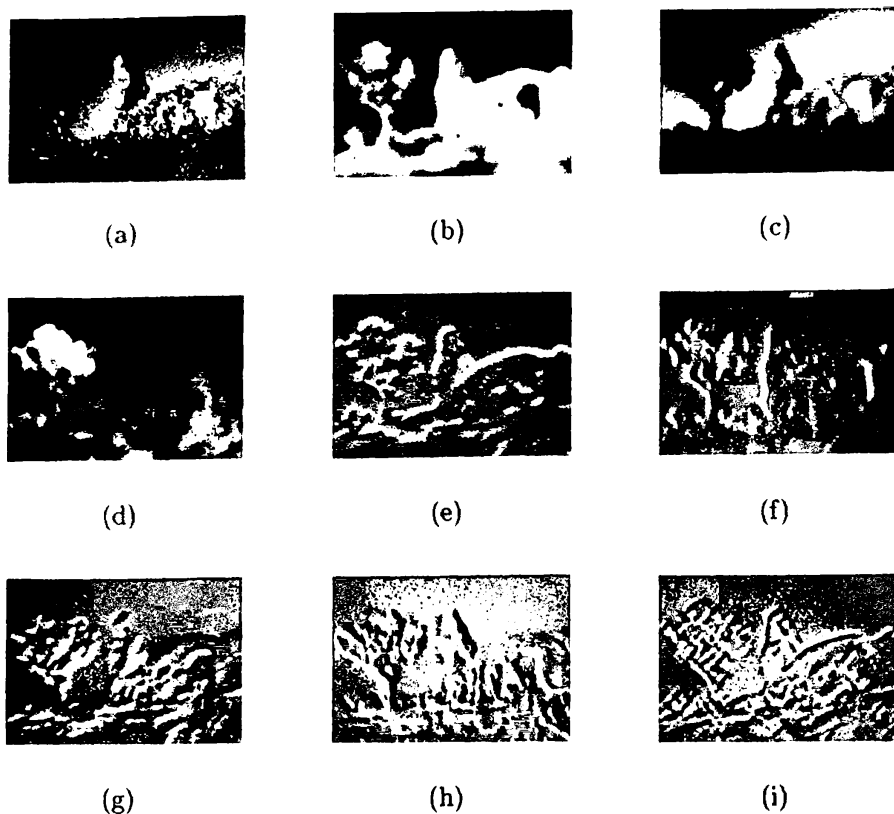
Στο σχήμα (2.1) δίνεται ένα παράδειγμα εικόνας και των MRF χαρακτηριστικών της (παρουσιάζονται οι εικόνες των πρώτων 8 MRF συνιστωσών μετά από εφαρμογή PCA (2.4) για μείωση διάστασης). Χρησιμοποιείται τετραγωνικό παράθυρο γειτονιάς 7×7 , που αφορά το κεντρικό εικονοστοιχείο x_c , και το οποίο χρησιμοποιείται και στα πειράματα της εργασίας. Οι εικόνες που χρησιμοποιούνται είναι έγχρωμες, επομένως το διάνυσμα τιμών του x_c ανήκει στον $7 * 7 * 3 = 147$ -διάστατο χώρο.

2.3 SIFT χαρακτηριστικά

Τα SIFT (Scale Invariant Feature Transform) χαρακτηριστικά [8], [7], [14] αποτελούν χαρακτηριστικά σημεία μίας εικόνας. Συγκεκριμένα, είναι αμετάβλητα στο χώρο κλίμακας της εικόνας και την περιστροφή και μερικώς αμετάβλητα σε αλλαγές φωτεινότητας και την τρισδιάστατη αντίληψη της κάμερας.

Στην παρουσία εργασίας μας ενδιαφέρει μία παραλλαγή των SIFT χαρακτηριστικών, όπου





Σχήμα 2.1: MRF χαρακτηριστικά. (a) Εικόνα; (b) 1η MRF συνιστώσα; (c) 2η MRF συνιστώσα; (d) 3η MRF συνιστώσα; (e) 4η MRF συνιστώσα; (f) 5η MRF συνιστώσα; (g) 6η MRF συνιστώσα; (h) 7η MRF συνιστώσα; (i) 8η MRF συνιστώσα;

δε χρησιμοποιούμε το βήμα εντοπισμού τους στην εικόνα αλλά υπολογίζουμε για κάθε εικονοστοιχείο της (είτε είναι σημείο ενδιαφέροντος είτε όχι) τις τοπικές κατευθύνσεις της κλίσης της παραγώγου της εικόνας. Μετά από αυτούς τους υπολογισμούς, προκύπτει ένας περιγραφέας για κάθε εικονοστοιχείο της εικόνας.

Επιθυμούμε να αναθέσουμε έναν προσανατολισμό σε κάθε σημείο της εικόνας, χρησιμοποιώντας την τοπική πληροφορία γύρω του. Για την εικόνα I υπολογίζουμε το μέτρο της παραγώγου m και τον προσανατολισμό θ σε κάθε σημείο της (x, y)

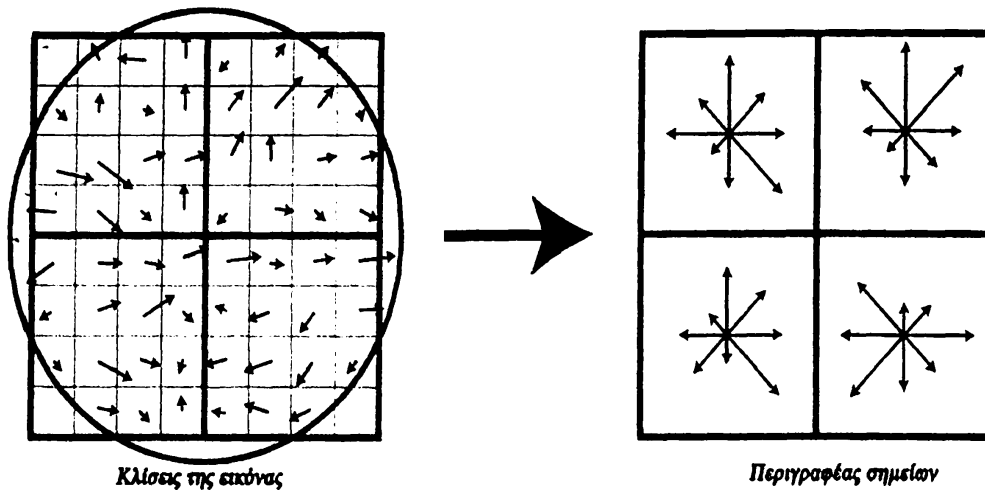
$$m(x, y) = \sqrt{(I(x+1, y) - I(x-1, y))^2 + (I(x, y+1) - I(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1} \left(\frac{I(x, y+1) - I(x, y-1)}{I(x+1, y) - I(x-1, y)} \right)$$

Ο περιγραφέας κάθε εικονοστοιχείου της εικόνας προκύπτει υπολογίζοντας το μέτρο και τον προσανατολισμό της παραγώγου σε περιοχή μεγέθους 16×16 γύρω από το εικονοστοιχείο. Στην περιοχή αυτή υπολογίζονται ιστογράμματα των κλίσεων για κάθε υποπεριοχή μεγέθους 4×4 , όπου για κάθε ιστογράμμο κρατάμε 8 κάδους κλίσεων. Συνεπώς έχουμε 16 ιστογράμματα με 8 κάδους το καθένα. Επομένως, ο περιγραφέας αποτελείται από ένα διάνυσμα που περιέχει τις τιμές όλων των ιστογραμμάτων, δηλαδή $16 \times 8 = 128$ τιμές.



Στο σχήμα (2.2) χρησιμοποιείται περιοχή μεγέθους 8×8 (αριστερή εικόνα) γύρω από το εικονοστοιχείο. Συνεπώς, έχουμε 4 ιστογράμματα με 8 κάδους το καθένα. Τα 4 ιστογράμματα απεικονίζονται ως πίνακας ιστογραμμάτων 2×2 (δεξιά εικόνα).



Σχήμα 2.2: Οι κλίσεις της εικόνας γύρω από το εικονοστοιχείο και ο περιγραφέας του. Το σχήμα αναπαράχθηκε από το [8].

2.4 Ανάλυση σε πρωτεύουσες συνιστώσες

Η τεχνική της ανάλυσης πρωτευουσών συνιστωσών (PCA) [1], [21] είναι ευρύτερα διαδεδομένη και χρησιμοποιείται για μείωση της διάστασης των δεδομένων, μεγιστοποιώντας παράλληλα τη διακύμανση των προβαλλόμενων δεδομένων. Με αυτόν τον τρόπο αντιμετωπίζεται το φαινόμενο της κατάρτας της διαστατικότητας, που σημαίνει ότι πολλά είδη ανάλυσης δεδομένων γίνονται σημαντικά δυσκολότερα όσο περισσότερο αυξάνεται η διάσταση των δεδομένων. Αυτό έχει ως αποτέλεσμα πολλοί αλγόριθμοι ταξινόμησης και ομαδοποίησης να έχουν μειωμένη απόδοση στην ακρίβεια της ταξινόμησης και την ποιότητα των ομάδων που παράγουν, αντίστοιχα.

2.4.1 Μαθηματικές λεπτομέρειες

Δοθέντος ενός πίνακα δεδομένων D μεγέθους $m \times n$, όπου οι m γραμμές είναι τα αντικείμενά μας και οι n στήλες είναι τα γνωρίσματά τους, ο πίνακας συμμεταβλητότητας του D είναι ο πίνακας S , ο οποίος ορίζεται

$$\forall i, j = 1, \dots, n : s_{ij} = \frac{\sum_{k=1}^m (d_{ki} - \bar{d}_{*i})(d_{kj} - \bar{d}_{*j})}{m} \quad (2.1)$$

όπου ο S είναι μεγέθους $n \times n$, d_{ki} είναι η k -οστή συνιστώσα της i -οστής στήλης, d_{kj} είναι η k -οστή συνιστώσα της j -οστής στήλης, $\overline{d_{*i}}$ είναι η μέση τιμή της i -οστής στήλης και $\overline{d_{*j}}$ είναι η μέση τιμή της j -οστής στήλης.

Στόχος μας είναι να προβάλλουμε τα δεδομένα σε ένα χώρο με διάσταση $M < n$, ενώ παράλληλα θέλουμε να μεγιστοποιείται η διακύμανση των προβαλλόμενων δεδομένων. Ας θεωρήσουμε την προβολή σε μία διάσταση ($M = 1$). Μπορούμε να ορίσουμε την κατεύθυνση αυτού του μονοδιάστατου χώρου χρησιμοποιώντας ένας n -διάστατο διάνυσμα u_1 το οποίο για ευκολία (και χωρίς βλάβη της γενικότητας) επιλέγουμε να είναι μοναδιαίου μήκους έτσι ώστε $u_1^T u_1 = 1$. Κάθε δεδομένο d_{i*} (θεωρούμε ως d_{i*} την i -οστή γραμμή του πίνακα D) προβάλλεται τότε στη βαθμωτή τιμή $u_1^T d_{i*}$. Η μέση τιμή των προβαλλόμενων δεδομένων είναι

$$\overline{d_{new}} = u_1^T \bar{d} \quad (2.2)$$

όπου \bar{d} είναι η απλή μέση τιμή $\bar{d} = \frac{1}{m} \sum_{k=1}^m d_{k*}$ και η διακύμανση των προβαλλόμενων δεδομένων είναι

$$S_{new} = \frac{1}{m} \sum_{k=1}^m \{u_1^T d_{k*} - u_1^T \bar{d}\}^2 = u_1^T S u_1 \quad (2.3)$$

όπου ο πίνακας συμμεταβλητότητας S δίνεται από τη σχέση (2.1).

Μεγιστοποιούμε την προβαλλόμενη συνδιακύμανση S_{new} ως προς το u_1 . Ωστόσο, θα πρέπει να γίνει μεγιστοποίηση με περιορισμό για να αποφύγουμε $\|u_1\| \rightarrow \infty$. Ο κατάλληλος περιορισμός προκύπτει από τη συνθήκη κανονικοποίησης $u_1^T u_1 = 1$. Για να επιβάλλουμε αυτόν τον περιορισμό, εισάγουμε τον πολλαπλασιαστή Lagrange που τον συμβολίζουμε ως λ_1 και έπειτα πραγματοποιούμε βελτιστοποίηση χωρίς περιορισμούς του: $u_1^T S u_1 + \lambda_1 (1 - u_1^T u_1)$. Θέτοντας την παράγωγο ως προς u_1 ίση με μηδέν, βλέπουμε ότι αυτή η ποσότητα έχει σταθερό σημείο όταν $S u_1 = \lambda_1 u_1$, το οποίο σημαίνει ότι το u_1 πρέπει να είναι ιδιοδιάνυσμα του S . Αν πολλαπλασιάσουμε από αριστερά με u_1^T και χρησιμοποιήσουμε ότι $u_1^T u_1 = 1$, βλέπουμε ότι η μεταβλητότητα δίνεται από: $u_1^T S u_1 = \lambda_1$. Αυτό το ιδιοδιάνυσμα είναι γνωστό ως η πρώτη κύρια συνιστώσα (principal component).

Μπορούμε να ορίσουμε επιπρόσθετες πρωτεύουσες συνιστώσες αυξητικά διαλέγοντας κάθε καινούρια κατεύθυνση να είναι τέτοια ώστε να μεγιστοποιεί την προβαλλόμενη διακύμανση ανάμεσα σε όλες τις πιθανές κατευθύνσεις, οι οποίες πρέπει να είναι ορθογόνιες ως προς αυτές που έχουν ήδη γίνει αποδεκτές. Αν θεωρήσουμε τη γενική περίπτωση ενός M -διάστατου χώρου προβολής, η βέλτιστη γραμμική προβολή για την οποία μεγιστοποιείται η διακύμανση των προβαλλόμενων δεδομένων καθορίζεται από τα M ιδιοδιανύσματα u_1, u_2, \dots, u_M του πίνακα συμμεταβλητότητας S , τα οποία αντιστοιχούν στις M μεγαλύτερες ιδιοτιμές $\lambda_1, \lambda_2, \dots, \lambda_M$. Αυτό μπορεί εύκολα ναδειχτεί χρησιμοποιώντας απόδειξη με επαγωγή.

Έστω $U = [u_1, u_2, \dots, u_M]$ είναι ο πίνακας με τα M μεγαλύτερα ιδιοδιανύσματα του S . Δηλαδή, το ιδιοδιάνυσμα u_i είναι το i -οστό μεγαλύτερο ιδιοδιάνυσμα. Επίσης, υποθέτουμε ότι ο πίνακας δεδομένων D έχει προεπεξεργαστεί έτσι ώστε η μέση τιμή του κάθε γνωρίσματος (στήλη) να είναι 0. Ισχύουν οι παρακάτω διαπιστώσεις:



- Ο νέος πίνακας δεδομένων $D' = DU$ είναι τα μετασχηματισμένα δεδομένα που ικανοποιούν τις εξής συνθήκες:
 - Κάθε ζεύγος νέων γνωρισμάτων έχει συμμεταβλητότητα 0.
 - Τα γνωρίσματα ταξινομούνται σύμφωνα με το πόση από τη διακύμανση των δεδομένων καταλαμβάνει κάθε νέο γνώρισμα.
 - Το πρώτο νέο γνώρισμα καταλαμβάνει όσο το δυνατόν περισσότερη από τη διακύμανση των δεδομένων.
 - Δεδομένης της απαίτησης της ορθογωνιότητας με τα προηγούμενα γνωρίσματα,
 - κάθε νέο γνώρισμα καταλαμβάνει όσο το δυνατόν περισσότερη από την υπολειπόμενη διακύμανση.
- Κάθε νέο γνώρισμα είναι ένας γραμμικός συνδυασμός των αρχικών γνωρισμάτων. Συγκεκριμένα, τα βάρη του γραμμικού συνδυασμού του i -οστού γνωρίσματος είναι οι συνιστώσες του i -οστού ιδιοδιανύσματος. Αυτό προκύπτει από το γεγονός ότι η j -οστή στήλη του D' δίνεται από το γινόμενο Du_j .
- Η διακύμανση του i -οστού νέου γνωρίσματος είναι ίση με λ_i .
- Το άθροισμα της διακύμανσης των αρχικών γνωρισμάτων είναι ίσο με το άθροισμα των διακυμάνσεων των νέων γνωρισμάτων αν κρατήσουμε όλα τα ιδιοδιανύσματα, δηλαδή αν $M = n$.
- Τα νέα γνωρίσματα καλούνται πρωτεύουσες συνιστώσες, δηλαδή το πρώτο νέο γνώρισμα καλείται πρώτη πρωτεύουσα συνιστώσα, το δεύτερο νέο γνώρισμα καλείται δεύτερη πρωτεύουσα συνιστώσα και ούτω καθεξής.

2.4.2 Διαπιστώσεις

Η ανάλυση πρωτευουσών συνιστωσών (PCA) μειώνει τη διάσταση των δεδομένων, μεγιστοποιώντας παράλληλα τη διακύμανση των προβαλλόμενων δεδομένων. Επομένως, εγγυάται την ελάχιστη δυνατή απώλεια πληροφορίας με την έννοια της μέγιστης κατά το δυνατόν διατήρησης της απόστασης των δεδομένων μετά την προβολή τους σε χώρο μικρότερης διάστασης. Συγκεκριμένα:

- Το ιδιοδιάνυσμα που σχετίζεται με τη μεγαλύτερη ιδιοτιμή υποδεικνύει την κατεύθυνση στην οποία τα δεδομένα έχουν την περισσότερη διακύμανση. Με άλλα λόγια, αν όλα τα διανύσματα δεδομένων προβληθούν πάνω στη γραμμή που ορίζεται από αυτό το ιδιοδιάνυσμα, τα διανύσματα που προκύπτουν θα έχουν τη μέγιστη διακύμανση σε σχέση με όλες τις πιθανές κατευθύνσεις.
- Το ιδιοδιάνυσμα που σχετίζεται με τη δεύτερη μεγαλύτερη ιδιοτιμή είναι η κατεύθυνση (ορθογώνια σε αυτή του πρώτου ιδιοδιανύσματος) στην οποία τα δεδομένα έχουν την περισσότερη υπολειπόμενη διακύμανση.



- Το πρώτο νέο γνώρισμα καταλαμβάνει όσο το δυνατόν περισσότερη από τη διακύμανση των δεδομένων.
- Δεδομένης της απαίτησης της ορθογωνιότητας με τα προηγούμενα γνωρίσματα, κάθε νέο γνώρισμα καταλαμβάνει όσο το δυνατόν περισσότερη από την υπολειπόμενη διακύμανση.
- Τα ιδιοδιανύσματα του S ορίζουν νέους άξονες. Πράγματι, η PCA μπορεί να θεωρηθεί σαν περιστροφή των αρχικών αξόνων συντεταγμένων σε νέους άξονες που είναι ευθυγραμμισμένοι με τη μεταβλητότητα στα δεδομένα. Η συνολική μεταβλητότητα των δεδομένων διατηρείται, αλλά τα νέα γνωρίσματα είναι τώρα ασυσχέτιστα.

2.5 Φασματική ομαδοποίηση σε επίπεδο εικονοστοιχείου

Στόχος της φασματικής ομαδοποίησης σε επίπεδο εικόνας είναι η ομαδοποίηση των εικονοστοιχείων της σε ομοιογενείς περιοχές οι οποίες ονομάζονται υπερεικονοστοιχεία. Η ομοιογένεια κάθε περιοχής αφορά χαρακτηριστικά και ιδιότητες, όπως απόσταση, χρώμα, φωτεινότητα, υφή κ.λ.π. Επιπλέον, θα πρέπει διαφορετικές περιοχές εικονοστοιχείων να είναι διακριτές μεταξύ τους όσον αφορά τα αντίστοιχά τους χαρακτηριστικά. Γενικότερα, η φασματική ομαδοποίηση αξιοποιεί τις ιδιότητες ενός γραφήματος (που κατασκευάζει), χρησιμοποιώντας την πληροφορία που μεταβιβάζεται από τη φασματική αποσύνθεση ενός πίνακα ομοιότητας. Τα στοιχεία αυτού του πίνακα κωδικοποιούν τις βαθύτερες ομοιότητες μεταξύ των κόμβων του γραφήματος (εδώ εικονοστοιχεία της εικόνας). Ακολουθεί η περιγραφή δύο μεθόδων που χρησιμοποιούνται ευρέως και υλοποιούν φασματική ομαδοποίηση, ακολουθώντας διαφορετική στρατηγική βελτιστοποίησης.

2.5.1 Φασματική ομαδοποίηση με χρήση του Λαπλασιανού πίνακα

Αρχικά θα επικεντρωθούμε στο απλούστερο έργο του διαχωρισμού ενός συνόλου δεδομένων, έστω X , σε δύο ομάδες A και B . Στο τέλος, θα αναφερθούν κάποιες γενικεύσεις. Έστω τα δεδομένα $X = \{x_1, x_2, \dots, x_N\}$. Τότε, πραγματοποιούμε τα παρακάτω βήματα:

- Κατασκευάζουμε ένα γράφημα $G(V, E)$, όπου κάθε κόμβος του γραφήματος αντιστοιχεί σε ένα σημείο $x_i, i = 1, 2, \dots, N$ του X . Επίσης, θεωρούμε ότι το G είναι μη κατευθυνόμενο και συνδεδεμένο.
- Θεωρούμε βάρος σε κάθε ακμή e_{ij} του γραφήματος G μέσω ενός πίνακα $W(i, j)$ που μετρά την ομοιότητα μεταξύ των αντίστοιχων κόμβων v_i, v_j . Το σκετ βαρών ορίζει την εγγύτητα (μερικές φορές καλείται συγγένεια) μέσω του $N \times N$ πίνακα W με στοιχεία

$$W \equiv [W(i, j)], \quad i, j = 1, 2, \dots, N$$

Ο πίνακας ομοιότητας θεωρείται ότι είναι συμμετρικός, δηλαδή $W(i, j) = W(j, i)$. Η επιλογή των βαρών αφήνεται στο χρήστη και αφορά κάθε πρόβλημα που διαπραγμα-



τεύεται [12]. Μία κοινή επιλογή είναι

$$W(i, j) = \begin{cases} e^{-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}} & \text{αν } \|x_i - x_j\|_2 < \epsilon \\ 0 & \text{διαφορετικά.} \end{cases}$$

όπου το σ και το ϵ είναι σταθερές καθορισμένες από το χρήστη. Η επιλογή του πίνακα ομοιότητας γενικά δεν είναι εύκολο έργο. Μία σωστή επιλογή μπορεί να έχει σημαντική βελτίωση στα αποτελέσματα που βρίσκουμε. Επίσης, οι σταθερές σ και ϵ θα πρέπει να καθοριστούν πειραματικά, ώστε τελικά ο πίνακας W να είναι πλήρως καθορισμένος βάσει των απαιτήσεων του πρόβληματος που θέλουμε να επιλύσουμε.

- Επιλέγουμε ένα κατάλληλο κριτήριο ομαδοποίησης για το διαχωρισμό του γράφου. Αρχικά ορίζουμε

$$cut(A, B) = \sum_{i \in A, j \in B} W(i, j) \quad (2.4)$$

$$D_{ii} = \sum_{j \in V} W(i, j), \quad i = 1, 2, \dots, N \quad (2.5)$$

$$V(A) = \sum_{i \in A} D_{ii} = \sum_{i \in A, j \in V} W(i, j) \quad (2.6)$$

Χρησιμοποιώντας τις (2.4), (2.5) και (2.6), το κριτήριο που χρησιμοποιείται εδώ είναι το $Ncut$ που ορίζεται

$$Ncut(A, B) = \frac{cut(A, B)}{V(A)} + \frac{cut(A, B)}{V(B)} \quad (2.7)$$

Ελαχιστοποιώντας το $Ncut(A, B)$, πετυχαίνουμε δύο ομάδες κόμβων του G που μεταξύ τους σχετίζονται ελάχιστα. Η ελαχιστοποίηση του $Ncut(A, B)$ ένα $NP - hard$ πρόβλημα. Για να παρακάμψουμε αυτή τη δυσκολία, θα επαναδιατυπώσουμε το πρόβλημα για να το φέρουμε σε μία μορφή που επιτρέπει μία αποδοτική προσεγγιστική λύση. Έστω λοιπόν

$$y_i = \begin{cases} \frac{1}{V(A)} & \text{αν } i \in A \\ -\frac{1}{V(B)} & \text{αν } i \in B \end{cases} \quad (2.8)$$

και θεωρούμε

$$y = [y_1, y_2, \dots, y_N]^T \quad (2.9)$$

Κάθε y_i μπορεί να θεωρηθεί ως δείκτης ομάδας του αντίστοιχου σημείου $x_i, i = 1, 2, \dots, N$. Τότε χρησιμοποιώντας τους ορισμούς των (2.8) και (2.9) ισχύει

$$\begin{aligned} y^T L y &= \frac{1}{2} \sum_{i \in V} \sum_{j \in V} (y_i - y_j)^2 W(i, j) \\ &= \sum_{i \in A} \sum_{j \in B} \left(\frac{1}{V(A)} + \frac{1}{V(B)} \right)^2 cut(A, B) \\ &\propto \left(\frac{1}{V(A)} + \frac{1}{V(B)} \right)^2 cut(A, B) \end{aligned} \quad (2.10)$$



καθώς η συνεισφορά του $y_i - y_j$ είναι μηδέν για τα σημεία στην ίδια ομάδα. Το σύμβολο α δηλώνει αναλογία και

$$L = D - W, \quad D \equiv \text{diag}[D_{ii}] \quad (2.11)$$

καλείται ο Λαπλασιανός πίνακας του γραφήματος. Ο πίνακας D είναι διαγώνιος με τα στοιχεία D_{ii} κατά μήκος της κύριας διαγωνίου. Επιπρόσθετα έχουμε

$$\begin{aligned} y^T D y &= \sum_{i \in A} y_i^2 D_{ii} + \sum_{j \in B} y_j^2 D_{jj} \\ &= \frac{1}{V(A)^2} V(A) + \frac{1}{V(B)^2} V(B) \\ &= \frac{1}{V(A)} + \frac{1}{V(B)} \end{aligned} \quad (2.12)$$

Επομένως, βάσει των (2.10) και (2.12), το πρόβλημα ελαχιστοποίησης του Ncut (2.7) είναι ισοδύναμο με την ελαχιστοποίηση του

$$J = \frac{y^T L y}{y^T D y} \quad (2.13)$$

με τον περιορισμό ότι $y_i \in \{\frac{1}{V(A)}, -\frac{1}{V(B)}\}$. Επιπλέον, με άμεση αντικατάσταση στους ορισμούς των αναμεμειγμένων ποσοτήτων έχουμε

$$y^T D \mathbf{1} = 0 \quad (2.14)$$

όπου $\mathbf{1}$ είναι ένα N -διάστατο διάνυσμα με όλες τις συνιστώσες του ίσες με 1. Προκειμένου να ξεπεράσουμε την $NP - hard$ φύση του προβλήματος, θα λύσουμε το απλουστευμένο πρόβλημα (2.13) υπό τον περιορισμό (2.14). Οι άγνωστες μεταβλητές $y_i, i = 1, 2, \dots, N$ υποθέτουμε ότι βρίσκονται στον άξονα των πραγματικών αριθμών. Είμαστε ήδη πολύ κοντά σε ένα γνωστό πρόβλημα βελτιστοποίησης. Ορίζουμε

$$z \equiv D^{1/2} y$$

Τότε η (2.13) γίνεται

$$J = \frac{z^T \bar{L} z}{z^T z} \quad (2.15)$$

και ο περιορισμός (2.14) γίνεται

$$z^T D^{1/2} \mathbf{1} = 0 \quad (2.16)$$

όπου $\bar{L} \equiv D^{-1/2} L D^{-1/2}$ και είναι γνωστός ως κανονικοποιημένος Λαπλασιανός πίνακας γραφήματος. Μπορεί εύκολα ναδειχτεί ότι ο \bar{L} έχει τις ακόλουθες ιδιότητες.

- Είναι συμμετρικός και μη αρνητικά ορισμένος. Επομένως, όλες οι ιδιοτιμές του είναι μη αρνητικές και τα αντίστοιχα ιδιοδιανύσματα του, που υπολογίζονται όπως στο [3], είναι ορθογώνια μεταξύ τους ανά δύο.



- Εύκολα μπορεί να ελεγχθεί ότι το $D^{1/2}\mathbf{1}$ είναι ιδιοδιάνυσμα που αντιστοιχεί στη μηδενική ιδιοτιμή, δηλαδή $\bar{L}D^{1/2}\mathbf{1} = 0$. Προφανώς, η μηδενική ιδιοτιμή είναι η μικρότερη του \bar{L} , αφού είναι μη αρνητικά ορισμένος.

Έχουμε επομένως όλα τα συστατικά για να πραγματοποιήσουμε την τελική βελτιστοποίηση. Παρατηρούμε ότι ο λόγος (2.15) είναι το γνωστό πηλίκο Rayleigh. Από τη γραμμική άλγεβρα, γνωρίζουμε ότι

- Η μικρότερη τιμή του πηλίκου, σε σχέση με το x , ισούται με τη μικρότερη ιδιοτιμή του \bar{L} και συμβαίνει για z ίσο με το ιδιοδιάνυσμα που αντιστοιχεί σε αυτή τη μικρότερη ιδιοτιμή.
- Αν περιορίσουμε τη λύση να είναι ορθογώνια σε όλα τα ιδιοδιανύσματα που σχετίζονται με τις j μικρότερες ιδιοτιμές $\lambda_0, \lambda_1, \dots, \lambda_{j-1}$, το πηλίκο Rayleigh ελαχιστοποιείται από το ιδιοδιάνυσμα που αντιστοιχεί στην επόμενη μικρότερη ιδιοτιμή λ_j και η ελάχιστη τιμή του ισούται με λ_j .

Λαμβάνοντας υπόψη τη συνθήκη ορθογωνιότητας στον περιορισμό (2.16) και το γεγονός ότι το $D^{1/2}\mathbf{1}$ είναι το ιδιοδιάνυσμα που αντιστοιχεί στη μικρότερη ιδιοτιμή $\lambda_0 = 0$, καταλήγουμε ότι το πρόβλημα ελαχιστοποίησης του πηλίκου Rayleigh (2.15) υπό τον περιορισμό (2.16), δηλαδή το

$$\min_z \left(\frac{z^T \bar{L} z}{z^T z} \right), \text{ με τον περιορισμό } z^T D^{1/2} \mathbf{1} = 0 \quad (2.17)$$

έχει βέλτιστη λύση που δίνεται από το ιδιοδιάνυσμα που αντιστοιχεί στη δεύτερη μικρότερη ιδιοτιμή του \bar{L} .

Για τη γενικότερη περίπτωση που θέλουμε να διαχωρίσουμε το γράφημα σε $m > 2$ περιοχές, θα πρέπει να υπολογίσουμε τα m ιδιοδιανύσματα που σχετίζονται με τις $\lambda_1, \lambda_2, \dots, \lambda_m$ ιδιοτιμές. Για αυτή την περίπτωση, οι περιορισμοί μας εμποδίζουν από το να απεικονίσουμε σε υποχώρο διάστασης μικρότερης από το επιθυμητό m . Για παράδειγμα, δε θέλουμε να προβάλλουμε σε ένα τριδιάστατο χώρο και τα σημεία να βρίσκονται σε ένα διδιάστατο επίπεδο ή μία μονοδιάστατη γραμμή.

- Τέλος εφαρμόζουμε έναν αποδοτικό αλγόριθμο για να πραγματοποιήσουμε το διαχωρισμό, σύμφωνα με το προηγούμενα υιοθετημένο κριτήριο ομαδοποίησης.

Παρακάτω ακολουθεί συνοπτικά η περιγραφή του αλγορίθμου (1) στο πλαίσιο της ομαδοποίησης των εικονοστοιχείων μιας εικόνας σε υπερεικονοστοιχεία.

Στο βήμα 1, ο πίνακας βαρών W κατασκευάζεται έτσι ώστε να έχει μεγάλη τιμή για εικονοστοιχεία που είναι όμοια μεταξύ τους και μικρή τιμή για εικονοστοιχεία που διαφέρουν αρκετά. Αυτό εξασφαλίζεται με τους δύο όρους του γινομένου. Ο πρώτος όρος εκφράζει ένα μέτρο ομοιότητας των εικονοστοιχείων όσον αφορά τα χαρακτηριστικά τους γνωρίσματα, ενώ ο δεύτερος όρος λαμβάνει υπόψη την χωρική απόσταση των εικονοστοιχείων. Οι συναρτήσεις F και X καθώς και το κατώφλι τ μπορούν να καθοριστούν πειραματικά. Στο βήμα 2, ο αριθμός M των μικρότερων ιδιοδιανυσμάτων που κρατάμε και σχηματίζουν



Αλγόριθμος 1 Φασματική ομαδοποίηση σε επίπεδο εικονοστοιχείου με χρήση του Λαπλασιανού πίνακα (Ncut)

1 Δοθείσης εικόνας I , κατασκευάζουμε το γράφημα $G(V, E)$ όπου κάθε κόμβος αντιστοιχεί σε ένα εικονοστοιχείο της και κάθε ακμή αφορά τη σύνδεση δύο εικονοστοιχείων. Θέτουμε το βάρος κάθε ακμής έτσι ώστε να καταδεικνύει το πόσο όμοια είναι τα εικονοστοιχεία που συνδέει. Συγκεκριμένα ο πίνακας βαρών W για $i, j \in V$ προκύπτει ως εξής:

$$w_{ij} = e^{-\frac{\|F(i)-F(j)\|_2^2}{\sigma_f^2}} * \begin{cases} e^{-\frac{\|X(i)-X(j)\|_2^2}{\sigma_x^2}} & \text{αν } \|X(i) - X(j)\|_2 < r \\ 0 & \text{διαφορετικά.} \end{cases}$$

όπου F συνάρτηση εξαγωγής χαρακτηριστικών γνωρισμάτων της εικόνας, X συνάρτηση απόστασης των εικονοστοιχείων της εικόνας και r τιμή κατωφλίου.

2 Υπολόγισε τις μικρότερες $M + 1$ ιδιοτιμές $\lambda_0, \lambda_1, \dots, \lambda_M$ του κανονικοποιημένου Λαπλασιανού πίνακα $\bar{L} \equiv D^{-1/2} L D^{-1/2}$ και τοποθέτησε τα αντίστοιχα ιδιοδιανύσματα, εκτός του πρώτου, u^1, \dots, u^M σε ένα πίνακα U ως στήλες. Ο U θα έχει την εξής μορφή:

$$U = \begin{bmatrix} u_1^1 & u_1^2 & \dots & u_1^M \\ u_2^1 & u_2^2 & \dots & u_2^M \\ \vdots & \vdots & \ddots & \vdots \\ u_n^1 & u_n^2 & \dots & u_n^M \end{bmatrix}$$

όπου M καθορίζεται από το χρήστη, n ο αριθμός των εικονοστοιχείων της εικόνας και u_i^j η i -οστή συνιστώσα του j -οστού μικρότερου ιδιοδιανύσματος του πίνακα βαρών W .

3 Θεωρώντας κάθε γραμμή i του πίνακα U ως διάνυσμα αναπαράστασης του εικονοστοιχείου i , εφαρμόζουμε τον αλγόριθμο K -μέσων (k -means) στις γραμμές του U και καθορίζουμε τον αριθμό ομάδων ίσο με M .

4 Κατατάσσουμε τα εικονοστοιχεία σε υπερεικονοστοιχεία σύμφωνα με την ομαδοποίηση που παράγεται από την εφαρμογή του αλγορίθμου K -μέσων.

τον πίνακα U συνήθως ταυτίζεται με τον αριθμό των ομάδων που θέλουμε να διαχωριστούν τα εικονοστοιχεία. Ωστόσο, οφείλουμε να διερευνήσουμε πειραματικά ποιός είναι ο θεωρητικά καλύτερος αριθμός ομάδων διαχωρισμού, ειδικά όταν εφαρμόζουμε τον αλγόριθμο Ncut [19], [18] σε ένα μεγάλο εύρος εικόνων (για παράδειγμα βάση εικόνων του Berkeley). Στο βήμα 3, εφαρμόζεται ο αλγόριθμος K -μέσων θεωρώντας ως δεδομένα ομαδοποίησης τις γραμμές του πίνακα U , δηλαδή σε κάθε εικονοστοιχείο i αντιστοιχίζεται το διάνυσμα της γραμμής i . Τέλος, στο βήμα 4 αναθέτουμε σε κάθε υπερεικονοστοιχείο εκείνα τα εικονοστοιχεία που κατέταξε στην ίδια κατηγορία ο αλγόριθμος K -μέσων.



Ο συγκεκριμένος αλγόριθμος βασίζεται στη λογική ότι η αντιληπτική ομαδοποίηση πρέπει να είναι μία διαδικασία που στοχεύει στο να εξάγουμε καθολικές εντυπώσεις από μία σκηνή και επιπλέον αποτελεί μία ιεραρχική περιγραφή της. Αντιμετωπίζοντας το πρόβλημα ομαδοποίησης ως πρόβλημα διαχωρισμού γράφου, επιτυγχάνουμε την ελαχιστοποίηση της συσχέτισης μεταξύ διαφορετικών ομάδων. Αυτή η λογική ακολουθείται στο [19].

2.5.2 Φασματική ομαδοποίηση με χρήση του πίνακα ομοιότητας

Ας υποθέσουμε ότι υπάρχουν k στοιχεία και c ομάδες. Μπορούμε να αναπαραστήσουμε μία ομάδα με ένα διάνυσμα με k στοιχεία. Επιτρέπουμε στα στοιχεία να σχετίζονται με τις ομάδες χρησιμοποιώντας κάποιο συνεχές βάρος. Πρέπει να είμαστε λίγο αόριστοι για τη σημασιολογία αυτών των βαρών, αλλά η πρόθεση είναι ότι αν η συνιστώσα σε ένα συγκεκριμένο διάνυσμα έχει μικρή τιμή, τότε σχετίζεται ασθενώς με την αντίστοιχη ομάδα, ενώ αν έχει μεγάλη τιμή, τότε σχετίζεται ισχυρά με την ομάδα.

Μία καλή ομάδα είναι αυτή στην οποία τα στοιχεία που συσχετίζονται ισχυρά με την ομάδα έχουν επίσης μεγάλες τιμές στη σύνδεσή τους στον πίνακα συγγένειας. Θεωρώντας A τον πίνακα που αναπαριστά τις συγγένειες μεταξύ των στοιχείων [12] και w_n το διάνυσμα βαρών σύνδεσης των στοιχείων με την n -οστή ομάδα, κατασκευάζουμε την εξής αντικειμενική συνάρτηση

$$J = w_n^T A w_n \quad (2.18)$$

Αυτό είναι ένα άθροισμα όρων της μορφής

$$\begin{aligned} [\text{συσχέτιση του στοιχείου } i \text{ με την ομάδα } n] &\times [\text{συγγένεια του στοιχείου } i \text{ με το στοιχείο } j] \\ &\times [\text{συσχέτιση του στοιχείου } j \text{ με την ομάδα } n] \end{aligned}$$

Μπορούμε να βρούμε μία ομάδα διαλέγοντας ένα σετ από βάρη συσχέτισης που μεγιστοποιούν την αντικειμενική συνάρτηση. Η αντικειμενική συνάρτηση είναι ανωφελής από μόνη της καθώς κλιμακώνοντας το w_n κατά λ κλιμακώνει τη συνολική συσχέτιση κατά λ^2 . Ωστόσο μπορούμε να κανονικοποιήσουμε τα βάρη απαιτώντας ότι

$$w_n^T w_n = 1 \quad (2.19)$$

Αυτό σημαίνει ότι πρέπει να μεγιστοποιήσουμε το (2.18) υπό τον περιορισμό (2.19). Πραγματοποιούμε βελτιστοποίηση με συντελεστές Lagrange και έχουμε το εξής πρόβλημα

$$\max(w_n^T A w_n + \lambda(w_n^T w_n - 1)) \quad (2.20)$$

όπου το λ είναι συντελεστής Lagrange. Με παραγωγή και αφαιρώντας τη σταθερά δύο παράγει το εξής

$$A w_n = \lambda w_n \quad (2.21)$$

που σημαίνει ότι το w_n είναι ιδιοδιάνυσμα του A . Αυτό σημαίνει ότι μπορούμε να σχηματίσουμε μία ομάδα αξιοποιώντας το ιδιοδιάνυσμα που αντιστοιχεί στη μεγαλύτερη ιδιοτιμή. Τότε, τα στοιχεία του ιδιοδιανύσματος είναι τα βάρη της ομάδας για κάθε στοιχείο. Για



προβλήματα που είναι εμφανείς κάποιες λογικές ομαδοποιήσεις, αναμένουμε ότι αυτά τα βάρη της ομάδας έχουν μεγάλη τιμή για μερικά στοιχεία, τα οποία ανήκουν στην ομάδα, και κοντά στο μηδέν για τα υπόλοιπα, τα οποία δεν ανήκουν στην ομάδα. Συγκεκριμένα, μπορούμε να πάρουμε τα βάρη για άλλες ομάδες από άλλα ιδιοδιανύσματα του A .

Σε χαρακτηριστικά προβλήματα όρασης, υπάρχουν ισχυρά βάρη συσχέτισης μεταξύ σχετικά λίγων ζευγών σημείων. Μπορούμε λογικά να αναμένουμε ότι ασχολούμαστε με ομάδες που είναι αρκετά σφιχτές και διακριτές.

Αυτές οι ιδιότητες οδηγούν σε μία αρκετά χαρακτηριστική δομή στον πίνακα συγγένειας. Συγκεκριμένα, αν μετονομάσουμε τους κόμβους του γραφήματος, τότε οι γραμμές και οι στήλες του πίνακα A αναμιγνύονται. Αναμένουμε να ασχολούμαστε με σχετικά μικρές συλλογές κόμβων με μεγάλα βάρη συσχέτισης. Επιπλέον, αναμένουμε ότι αυτές οι συλλογές πραγματικά σχηματίζουν μία σειρά από σχετικά συναφείς μεγαλοπρεπώς ασυνεχείς ομάδες. Αυτό σημαίνει ότι μπορούμε να αναμιξουμε τις γραμμές και τις στήλες του πίνακα δεδομένων A για να σχηματιστεί ένας πίνακας που είναι πρόχειρα μπλοκ-διαγώνιος (με τα μπλοκ του να είναι οι ομάδες). Αναμιγνύοντας τον A απλά ανακατεύει τα στοιχεία των ιδιοδιανυσμάτων του έτσι ώστε να μπορούμε να βγάλουμε συμπέρασμα σχετικά με τα ιδιοδιανύσματα σχεπτόμενοι μία ανακατεμένη έκδοση του A .

Τα ιδιοδιανύσματα μπλοκ-διαγώνιων πινάκων αποτελούνται από τα ιδιοδιανύσματα των μπλοκ γεμισμένα στο τέλος με μηδενικά. Αναμένουμε ότι κάθε μπλοκ έχει ένα ιδιοδιάνυσμα που αντιστοιχεί σε μία αρκετά υψηλή ιδιοτιμή (που αφορά τη συγκεκριμένη ομάδα) και έπειτα μία σειρά από μικρές ιδιοτιμές μη συγκεκριμένης σημασίας. Από αυτό, αναμένουμε ότι αν υπάρχουν c σημαντικές ομάδες (όπου $c < k$), τα ιδιοδιανύσματα που αντιστοιχούν στις c μεγαλύτερες ιδιοτιμές αναπαριστούν το καθένα μία ομάδα.

Αυτό σημαίνει ότι καθένα από αυτά τα ιδιοδιανύσματα είναι ένα ιδιοδιάνυσμα ενός μπλοκ γεμισμένο στο τέλος με μηδενικά. Συγκεκριμένα, ένα χαρακτηριστικό ιδιοδιάνυσμα έχει ένα μικρό σετ μεγάλων τιμών (που αντιστοιχούν στο μπλοκ) και ένα σετ από τιμές κοντά στο μηδέν. Αναμένουμε ότι μόνο ένα από αυτά τα ιδιοδιανύσματα έχει μεγάλη τιμή για οποιαδήποτε δοσμένη συνιστώσα, ενώ όλα τα άλλα θα είναι μικρά. Επομένως, μπορούμε να ερμηνεύσουμε τα ιδιοδιανύσματα που αντιστοιχούν στις c μεγαλύτερες κατά μέτρο ιδιοτιμές ως βάρη των ομάδων για τις πρώτες c ομάδες. Κάποιος μπορεί συνήθως να κβαντοποιεί τα βάρη των ομάδων στο μηδέν ή τη μονάδα για να εξασφαλίσει διακριτές ομάδες.

Αυτό είναι ένα ποιοτικό επιχείρημα και υπάρχουν γραφήματα για τα οποία το επιχείρημα είναι αναμφισβήτητο ύποπτο. Επιπλέον, ήμασταν ασαφείς σχετικά με το πώς να προσδιορίσουμε το c , παρόλο που το επιχείρημα μας προτείνει ότι η διερεύνηση κατά μήκος του φάσματος του A μπορεί να επιφέρει ορθότερα αποτελέσματα.

Παρακάτω ακολουθεί συνοπτικά η περιγραφή του αλγορίθμου (2) στο πλαίσιο της ομαδοποίησης των εικονοστοιχείων μιας εικόνας σε υπερεικονοστοιχεία.

Τα βήματα του παραπάνω αλγορίθμου έχουν επεξηγηθεί αναλυτικά στο τέλος της παραγράφου (2.5.1) μετά την περιγραφή του αλγορίθμου (1). Η λογική είναι παρόμοια με τη διαφορά ότι ακολουθείται διαφορετική στρατηγική βελτιστοποίησης. Τελικά επιτυγχάνουμε και πάλι καλά αποτελέσματα ομαδοποίησης εικονοστοιχείων.



Αλγόριθμος 2 Φασματική ομαδοποίηση σε επίπεδο εικονοστοιχείου με χρήση του πίνακα ομοιότητας

- 1 Δοθείσης εικόνας I , κατασκευάζουμε το γράφημα $G(V, E)$ όπου κάθε κόμβος αντιστοιχεί σε ένα εικονοστοιχείο της και κάθε ακμή αφορά τη σύνδεση δύο εικονοστοιχείων. Θέτουμε το βάρος κάθε ακμής έτσι ώστε να καταδεικνύει το πόσο όμοια είναι τα εικονοστοιχεία που συνδέει. Συγκεκριμένα ο πίνακας βαρών W για $i, j \in V$ προκύπτει ως εξής:

$$w_{ij} = e^{-\frac{\|F(i)-F(j)\|_2^2}{\sigma_f^2}} * \begin{cases} e^{-\frac{\|X(i)-X(j)\|_2^2}{\sigma_x^2}} & \text{αν } \|X(i) - X(j)\|_2 < r \\ 0 & \text{διαφορετικά.} \end{cases}$$

όπου F συνάρτηση εξαγωγής χαρακτηριστικών γνωρισμάτων της εικόνας, X συνάρτηση απόστασης των εικονοστοιχείων της εικόνας και r τιμή κατωφλίου.

- 2 Υπολόγισε τις μεγαλύτερες M ιδιοτιμές του πίνακα βαρών W και τοποθέτησε τα αντίστοιχα ιδιοδιανύσματα σε ένα πίνακα U ως στήλες. Ο U θα έχει την εξής μορφή:

$$U = \begin{bmatrix} u_1^1 & u_1^2 & \dots & u_1^M \\ u_2^1 & u_2^2 & \dots & u_2^M \\ \vdots & \vdots & \ddots & \vdots \\ u_n^1 & u_n^2 & \dots & u_n^M \end{bmatrix}$$

όπου M καθορίζεται από το χρήστη, n ο αριθμός των εικονοστοιχείων της εικόνας και u_i^j η i -οστή συνιστώσα του j -οστού μεγαλύτερου ιδιοδιανύσματος του πίνακα βαρών W .

- 3 Θεωρώντας κάθε γραμμή i του πίνακα U ως διάνυσμα αναπαράστασης του εικονοστοιχείου i , εφαρμόζουμε τον αλγόριθμο K -μέσων (k -means) στις γραμμές του U και καθορίζουμε τον αριθμό ομάδων ίσο με M .
- 4 Κατατάσσουμε τα εικονοστοιχεία σε υπερεικονοστοιχεία σύμφωνα με την ομαδοποίηση που παράγεται από την εφαρμογή του αλγορίθμου K -μέσων.
-

2.5.3 Παραδείγματα υπερεικονοστοιχείων

Παρακάτω στο σχήμα (2.3) παρατίθενται ορισμένα παραδείγματα εικόνων και τα υπερεικονοστοιχεία τους.

2.6 Φασματική ομαδοποίηση σε επίπεδο υπερεικονοστοιχείου

Στόχος του αλγορίθμου φασματικής ομαδοποίησης σε επίπεδο υπερεικονοστοιχείου [15], [17] είναι η ομαδοποίηση των υπερεικονοστοιχείων μιας εικόνας σε ευρύτερα ομοιογενείς





Σχήμα 2.3: Εικόνες και υπερακονοστοιχεία.

περιοχές. Η ομοιογένεια κάθε ευρύτερης περιοχής αφορά χαρακτηριστικά και ιδιότητες επιμέρους υπερακονοστοιχείων, όπως απόσταση, χρώμα, φωτεινότητα, υψή κ.λ.π.

Ο συγκεκριμένος αλγόριθμος θα μπορούσαμε να πούμε ότι ταυτίζεται με τον αλγόριθμο της ομαδοποίησης με χρήση του Διαπλαστικού Πίνακα, ο οποίος παρουσιάζεται στην ενότητα (2.5.1), αλλά τον εφαρμόζουμε σε επίπεδο υπερακονοστοιχείου. Σε περίπτωση, ο πίνακας βαρών W αφορά το πόσο παρόμοια είναι τα υπερακονοστοιχεία μεταξύ τους. Παρακάτω ακολουθεί η περιγραφή του αλγορίθμου (3).

Στο βήμα 1, ακολουθεί η διαδικασία που περιγράφεται στην ενότητα (2.5.1) εφαρμογή του αλγορίθμου Ncut στην εικόνα I . Ο αριθμός των υπερακονοστοιχείων της εικόνας καθορίζεται από το χρήστη, ανάλογα με το βαθμό ομοιογένειας που επιθυμεί να απαιτήσει το πρόβλημά του. Συγκεκριμένα, όσο μεγαλύτερο πλήθος υπερακονοστοιχείων επιθυμεί να υπολογιστούν, τόσο καλύτερη ομοιογένεια επιτυγχάνεται στο περιεχόμενο των υπερακονοστοιχείων της κατετιμημένης εικόνας, ωστόσο απαιτείται μεγαλύτερη χρονική διάρκεια και μνήμη, για τον υπολογισμό των υπερακονοστοιχείων, κάτι τ



Αλγόριθμος 3 Φασματική ομαδοποίηση σε επίπεδο υπερεικονοστοιχείου με χρήση του Λαπλασιανού πίνακα

- 1 Δοθείσης εικόνας I , εφαρμόζουμε τον αλγόριθμο Ncut για να βρούμε ένα δοσμένο αριθμό υπερεικονοστοιχείων, έστω K , της εικόνας I .
 - 2 Εφαρμόζουμε ξανά τον αλγόριθμο Ncut, όπου τώρα ο πίνακας ομοιότητας W αφορά το πόσο παραπλήσια είναι τα υπερεικονοστοιχεία μεταξύ τους.
-

αποτελεί δυσμενή παράγοντα σε εφαρμογές που αναμένουν άμεσα και ολοκληρωμένα αποτελέσματα. Στο βήμα 2, επαναλαμβάνουμε τη διαδικασία που περιγράφηκε στην ενότητα (2.5.1) με την εφαρμογή του αλγορίθμου Ncut στην εικόνα I αλλά αντιστοιχίζοντας κάθε κόμβο του γραφήματος σε ένα υπερεικονοστοιχείο, αντί για εικονοστοιχείο. Οπότε πλέον μας ενδιαφέρει πόσο "απέχουν" τα υπερεικονοστοιχεία μεταξύ τους για να ορίσουμε τον πίνακα βαρών W .

Μπορούμε να χρησιμοποιήσουμε, αντίστοιχα, και τον αλγόριθμο που παρουσιάστηκε στην ενότητα (2.5.2), αλλά εφαρμόζοντας τον σε επίπεδο υπερεικονοστοιχείου. Και πάλι σε αυτή την περίπτωση, ο πίνακας βαρών W αφορά το πόσο παραπλήσια είναι τα υπερεικονοστοιχεία μεταξύ τους. Ακολουθεί η περιγραφή του αλγορίθμου (4).

Αλγόριθμος 4 Φασματική ομαδοποίηση σε επίπεδο υπερεικονοστοιχείου με χρήση του πίνακα ομοιότητας

- 1 Δοθείσης εικόνας I , εφαρμόζουμε τον αλγόριθμο Ncut για να βρούμε ένα δοσμένο αριθμό υπερεικονοστοιχείων, έστω K , της εικόνας I .
 - 2 Εφαρμόζουμε τον αλγόριθμο φασματικής ομαδοποίησης με χρήση του πίνακα ομοιότητας, όπου τώρα ο πίνακας ομοιότητας W αφορά το πόσο παραπλήσια είναι τα υπερεικονοστοιχεία μεταξύ τους.
-

Τα βήματα του παραπάνω αλγορίθμου έχουν επεξηγηθεί αναλυτικά στο τέλος της παραγράφου αντίστοιχα μετά την περιγραφή του αλγορίθμου (3). Η λογική είναι παρόμοια με τη διαφορά ότι ακολουθείται διαφορετική στρατηγική βελτιστοποίησης. Τελικά επιτυγχάνουμε και πάλι καλά αποτελέσματα ομαδοποίησης υπερεικονοστοιχείων.

Στο κεφάλαιο 3 θα παρουσιαστούν αναλυτικά πειραματικά αποτελέσματα και για τους δύο αλγορίθμους ομαδοποίησης υπερεικονοστοιχείων.



2.7 Ταξινομητές

2.7.1 Απόκλιση Kullback–Leibler

Στη θεωρία πιθανοτήτων και θεωρία πληροφοριών, η απόκλιση Kullback–Leibler (KL) είναι γνωστή και ως απόκλιση πληροφορίας ή συλλογή πληροφοριών ή σχετική εντροπία. Αποτελεί ένα μη συμμετρικό μέτρο της διαφοράς μεταξύ δύο κατανομών πιθανότητας, έστω P και Q . Η KL μέτρα τον αναμενόμενο αριθμό των επιπλέον bits που απαιτούνται για να κωδικοποιηθούν δείγματα από την P κατά τη χρήση ενός κωδικού που βασίζεται στην Q , παρά τη χρησιμοποίηση ενός κώδικα που βασίζεται στην P . Συνήθως η P αντιπροσωπεύει την "πραγματική" κατανομή των δεδομένων, παρατηρήσεων ή την ακριβώς υπολογισμένη θεωρητική κατανομή. Το μέτρο Q αντιπροσωπεύει συνήθως μια θεωρία, ένα μοντέλο, μια περιγραφή ή την προσέγγιση της P .

Αν και συχνά θεωρείται ως μετρική απόσταση, η απόκλιση KL δεν είναι μια πραγματική μετρική (δεν είναι συμμετρική: η KL από την P στην Q δεν είναι απαραίτητα ίδια με την KL από την Q στην P).

Για τις κατανομές πιθανότητας P και Q μιας συνεχούς τυχαίας μεταβλητής, η KL απόκλιση ορίζεται ως το ολοκλήρωμα

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx \quad (2.22)$$

όπου p και q δηλώνουν τις πυκνότητες πιθανότητας των P και Q , αντίστοιχα. Στην παρούσα εργασία χρησιμοποιούμε τη σχέση (2.22) θεωρώντας ότι οι p και q είναι πυκνότητες πιθανότητες γκαουσιανής κατανομής, δηλαδή

$$p(x|\mu_1, \Sigma_1) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_1|^{1/2}} \exp \left\{ -\frac{1}{2}(x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) \right\} \quad (2.23)$$

$$q(x|\mu_2, \Sigma_2) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_2|^{1/2}} \exp \left\{ -\frac{1}{2}(x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2) \right\} \quad (2.24)$$

όπου μ είναι ένα D -διάστατο διάνυσμα μέσης τιμής, Σ είναι ένας $D \times D$ πίνακας συνδιακύμανσης και $|\Sigma|$ συμβολίζει την ορίζουσα του πίνακα Σ . Τότε αποδεικνύεται με αντικατάσταση ότι η KL απόκλιση [24] δίνεται από

$$D_{KL}(N(\mu_1, \Sigma_1), N(\mu_2, \Sigma_2)) = (\mu_1 - \mu_2)^T (\mu_1 - \mu_2) + Tr \left(\Sigma_1 + \Sigma_2 - 2(\Sigma_1 \Sigma_2)^{\frac{1}{2}} \right) \quad (2.25)$$

όπου μ_1, μ_2 είναι D -διάστατα διανύσματα μέσης τιμής, Σ_1, Σ_2 είναι $D \times D$ πίνακες συνδιακύμανσης και Tr συμβολίζει το άθροισμα των διαγώνιων στοιχείων του ορίσματος του.

Άλλη μία προσέγγιση της (2.25) δίνεται αν χρησιμοποιήσουμε μόνο τα διανύσματα μέσης τιμής, δηλαδή

$$D_{1-NN}(\mu_1, \mu_2) = (\mu_1 - \mu_2)^T (\mu_1 - \mu_2) \quad (2.26)$$

όπου τότε δεν πρόκειται για την KL απόκλιση, αλλά μπορεί να θεωρηθεί απόσταση τύπου χοντινότερου γείτονα.



2.7.2 Μηχανές διανυσμάτων υποστήριξης

Οι μηχανές διανυσμάτων υποστήριξης (SVMs) [21] αποτελούν μία τεχνική ταξινόμησης που έχει τις ρίζες της στη στατιστική θεωρία μάθησης και έχει αποδειχτεί ότι μπορεί να επιφέρει σημαντικά υποσχόμενα αποτελέσματα σε πολλές πρακτικές εφαρμογές από αναγνώριση χειρόγραφου κειμένου έως κατηγοριοποίηση κειμένου. Επίσης, δουλεύει πολύ καλά με υψηλής διάστασης δεδομένα και αποφεύγει το πρόβλημα της κατάρας της διαστατικότητας. Ακόμα μία μοναδική πτυχή αυτής της προσέγγισης είναι ότι αναπαριστά το σύνορο απόφασης χρησιμοποιώντας ένα υποσύνολο των παραδειγμάτων μάθησης, που είναι γνωστά ως διανύσματα υποστήριξης.

Η βασική ιδέα της τεχνικής που χρησιμοποιούν οι SVMs είναι ότι προσπαθούν να βρουν ένα υπερεπίπεδο το οποίο θα πρέπει να διαχωρίζει επαρκώς τα παραδείγματα μάθησης και, επιπλέον, να είναι κατάλληλο ώστε να μπορεί να πραγματοποιηθεί ορθή κατηγοριοποίηση άγνωστων παραδειγμάτων ελέγχου. Επομένως, είναι επιθυμητό να σχεδιάζονται SVMs που να μεγιστοποιούν τα περιθώρια του συνόρου απόφασης τους προκειμένου να εξασφαλίζουν ότι ελαχιστοποιούνται τα χειρότερης περίπτωσης σφάλματα γενίκευσης στα παραδείγματα ελέγχου. Ακολουθεί συνοπτική περιγραφή των διαφορετικών ταξινομητών SVMs που κατασκευάζονται ανάλογα με τα δεδομένα που θέλουμε να διαχωρίσουμε:

- *Γραμμικές SVMs: Διαχωρισμα δεδομένα*

Μία γραμμική SVM είναι ένας ταξινομητής που ψάχνει για ένα υπερεπίπεδο με το μεγαλύτερο περιθώριο και είναι γνωστός ως ταξινομητής μέγιστου περιθωρίου.

Έστω ένα δυαδικό πρόβλημα ταξινόμησης που αποτελείται από N παραδείγματα μάθησης. Κάθε παράδειγμα συμβολίζεται με το ζεύγος (x_i, y_i) για $i = 1, 2, \dots, N$, όπου $x_i = (x_{i1}, x_{i2}, \dots, x_{iN})$ αντιστοιχεί στο σύνολο γνωρισμάτων του i -οστού παραδείγματος. Θεωρούμε ως σύμβαση ότι $y_i \in \{-1, 1\}$, όπου το y_i δηλώνει την ετικέτα της κλάσης του x_i . Το σύνορο απόφασης ενός γραμμικού ταξινομητή μπορεί να γραφεί ως

$$w \cdot x + b = 0 \quad (2.27)$$

όπου w και b είναι παράμετροι του μοντέλου. Για παράδειγμα, αν τα σημεία x_a και x_b βρίσκονται πάνω στο σύνορο απόφασης, τότε πρέπει

$$w \cdot x_a + b = 0$$

$$w \cdot x_b + b = 0$$

Αφαιρώντας την πρώτη σχέση από τη δεύτερη προκύπτει

$$w \cdot (x_b - x_a) = 0$$

όπου το $x_b - x_a$ είναι ένα διάνυσμα παράλληλο στο σύνορο απόφασης και κατευθύνεται από το x_a στο x_b . Αφού το εσωτερικό γινόμενο είναι μηδέν, η κατεύθυνση του w θα πρέπει να είναι κάθετη στο σύνορο απόφασης.



Για οποιοδήποτε σημείο, έστω x_a , πάνω από το σύνορο απόφασης, μπορούμε να δείξουμε ότι

$$w \cdot x_a + b = k \quad (2.28)$$

όπου $k > 0$. Παρομοίως, για οποιοδήποτε σημείο, έστω x_c , κάτω από το σύνορο απόφασης, μπορούμε να δείξουμε ότι

$$w \cdot x_c + b = k' \quad (2.29)$$

όπου $k' < 0$. Αν βάλουμε ετικέτα κλάσης $+1$ σε όλα τα σημεία τύπου x_a και ετικέτα κλάσης -1 σε όλα τα σημεία τύπου x_c , τότε μπορούμε να προβλέψουμε την ετικέτα κλάσης, έστω y , για οποιοδήποτε παράδειγμα ελέγχου, έστω z , με τον ακόλουθο τρόπο

$$y = \begin{cases} 1 & \text{αν } w \cdot z + b > 0 \\ -1 & \text{αν } w \cdot z + b < 0 \end{cases} \quad (2.30)$$

Αν θεωρήσουμε το κοντινότερο σημείο στο σύνορο απόφασης τύπου x_a και το αντίστοιχο κοντινότερο σημείο τύπου x_c , τότε πρέπει να ικανοποιούν τις (2.28) και (2.29), αντίστοιχα. Μπορούμε να επανακλιμακώσουμε τις παραμέτρους w και b του συνόρου απόφασης έτσι ώστε τα δύο παράλληλα υπερεπίπεδα b_{i1} και b_{i2} να μπορούν να εκφραστούν ως εξής, αντίστοιχα

$$b_{i1} : w \cdot x + b = 1 \quad (2.31)$$

$$b_{i2} : w \cdot x + b = -1 \quad (2.32)$$

Το περιθώριο του συνόρου απόφασης δίνεται από την απόσταση μεταξύ των δύο αυτών υπερεπιπέδων. Για να υπολογίσουμε το περιθώριο, έστω x_1 ένα σημείο δεδομένων τοποθετημένο στο b_{i1} και x_2 ένα σημείο δεδομένων τοποθετημένο στο b_{i2} . Αντικαθιστώντας τα σημεία αυτά στις (2.31) και (2.32) και αφαιρώντας τη (2.32) από τη (2.31), το περιθώριο, έστω d , υπολογίζεται ως

$$w \cdot (x_1 - x_2) = 2$$

$$\|w\| \|x_1 - x_2\| \cos \theta = 2$$

$$d = \frac{2}{\|w\|} \quad (2.33)$$

όπου θ είναι η γωνία που σχηματίζουν τα διανύσματα w και $(x_1 - x_2)$.

Η εκπαίδευση της SVM περιλαμβάνει την εκτίμηση των παραμέτρων w και b του συνόρου απόφασης από τα παραδείγματα εκπαίδευσης. Οι παράμετροι πρέπει να επιλεγούν με τέτοιο τρόπο ώστε να επαληθεύονται οι ακόλουθες δύο συνθήκες

$$\begin{aligned} w \cdot x_i + b &\geq 1 & \text{αν } y_i &= 1 \\ w \cdot x_i + b &\leq -1 & \text{αν } y_i &= -1 \end{aligned} \quad (2.34)$$



Οι δύο ανισότητες της (2.34) μπορούν να συνοψιστούν σε μία πιο συμπαγή μορφή ως εξής

$$y_i(w \cdot x_i + b) \geq 1, \quad i = 1, 2, \dots, N. \quad (2.35)$$

Αν και οι προηγούμενες συνθήκες εφαρμόζονται επίσης σε οποιοδήποτε γραμμικό ταξινομητή, η SVM επιβάλλει μία επιπρόσθετη απαίτηση ότι το περιθώριο του συνόρου απόφασης πρέπει να είναι μέγιστο. Μεγιστοποιώντας το περιθώριο, ωστόσο, είναι ισοδύναμο με την ελαχιστοποίηση της ακόλουθης αντικειμενικής συνάρτησης

$$f(w) = \frac{\|w\|^2}{2} \quad (2.36)$$

Σημείωση 2.1 (Γραμμική SVM: Διαχωρισμα δεδομένα). Η διαδικασία μάθησης της SVM μπορεί να διατυπωθεί ως το ακόλουθο πρόβλημα βελτιστοποίησης με περιορισμούς

$$\min_w \frac{\|w\|^2}{2}$$

με τον περιορισμό ότι $y_i(w \cdot x_i + b) \geq 1, \quad i = 1, 2, \dots, N$

Καθώς η αντικειμενική συνάρτηση είναι τετραγωνική και οι περιορισμοί είναι γραμμικοί στις παραμέτρους w και b , αυτό είναι γνωστό ως κυρτό πρόβλημα βελτιστοποίησης, το οποίο μπορεί να λυθεί χρησιμοποιώντας την καθιερωμένη μέθοδο των πολλαπλασιαστών Lagrange.

- **Γραμμικές SVMs: Μη Διαχωρισμα δεδομένα**

Στην περίπτωση που τα δεδομένα δεν είναι εμφανώς διαχωρίσιμα, η γραμμική SVM πρέπει να τροποποιηθεί ώστε να μάθει ένα σύνορο απόφασης που να είναι ανεκτό σε μικρά λάθη εκπαίδευσης χρησιμοποιώντας μία μέθοδο που είναι γνωστή ως προσέγγιση χαλαρού περιθωρίου. Συγκεκριμένα, η μέθοδος που θα παρουσιαστεί επιτρέπει στην SVM να κατασκευάσει ένα γραμμικό σύνορο απόφασης ακόμα και σε περιπτώσεις όπου οι κλάσεις δεν είναι γραμμικά διαχωρίσιμες. Για να το κάνει αυτό, ο αλγόριθμος μάθησης της SVM θα πρέπει να λάβει υπόψη το ισοζύγιο ανάμεσα στο πλάτος του περιθωρίου και στον αριθμό των λαθών εκπαίδευσης που πραγματοποιούνται από το γραμμικό σύνορο απόφασης.

Αν και η αρχική αντικειμενική συνάρτηση της (2.36) είναι ακόμα εφαρμόσιμη, το σύνορο απόφασης που είχε βρεθεί δε συνεχίζει να ικανοποιεί όλους τους περιορισμούς της (2.35). Οι περιορισμοί ανισότητας πρέπει, επομένως, να χαλαρώσουν για να συμπεριλάβουμε τα μη γραμμικά διαχωρίσιμα δεδομένα. Αυτό μπορεί να γίνει εισάγοντας θετικών τιμών μεταβλητές χαλάρωσης (έστω ξ) στους περιορισμούς του προβλήματος βελτιστοποίησης ως εξής

$$\begin{aligned} w \cdot x_i + b &\geq 1 - \xi_i & \text{αν } y_i &= 1 \\ w \cdot x_i + b &\leq -1 + \xi_i & \text{αν } y_i &= -1 \end{aligned} \quad (2.37)$$

όπου $\forall i : \xi_i > 0$.



Η σημασία των μεταβλητών χαλάρωσης ξ_i είναι ότι μας προμηθεύουν μία εκτίμηση του λάθους του συνόρου απόφασης όσον αφορά το παράδειγμα μάθησης x_i .

Κατ' αρχήν μπορούμε να εφαρμόσουμε την ίδια αντικειμενική συνάρτηση όπως πριν και να επιβάλλουμε τις συνθήκες που δίνονται στην (2.37) προκειμένου να βρούμε το σύνορο απόφασης. Ωστόσο, καθώς δεν υπάρχουν περιορισμοί στον αριθμό των λαθών που μπορεί να κάνει το σύνορο απόφασης, ο αλγόριθμος μάθησης μπορεί να βρει ένα σύνορο απόφασης με πολύ μεγάλο περιώριο αλλά να ταξινομεί εσφαλμένα πολλά από τα παραδείγματα μάθησης. Για να αποφύγουμε αυτό το πρόβλημα, η αντικειμενική συνάρτηση πρέπει να τροποποιηθεί για να τιμωρεί ένα σύνορο απόφασης με μεγάλες τιμές στις μεταβλητές χαλάρωσης ξ_i . Η τροποποιημένη αντικειμενική συνάρτηση δίνεται από την ακόλουθη εξίσωση

$$f(w) = \frac{\|w\|^2}{2} + C \left(\sum_{i=1}^N \xi_i \right)^k \quad (2.38)$$

όπου C και k είναι παράμετροι καθορισμένες από το χρήστη και αναπαριστούν την τιμωρία της εσφαλμένης ταξινόμησης των παραδειγμάτων μάθησης. Μία στρατηγική επιλογής των κατάλληλων παραμέτρων C και k θα ήταν η πραγματοποίηση της τεχνικής cross-validation στο σύνολο των παραδειγμάτων μάθησης, όπου τελικά θα επιλέγαμε τις τιμές που μεγιστοποιούν την απόδοση του μοντέλου.

Σημείωση 2.2 (Γραμμική SVM: Μη Διαχωρίσιμα δεδομένα). Η διαδικασία μάθησης της SVM μπορεί να διατυπωθεί ως το ακόλουθο πρόβλημα βελτιστοποίησης με περιορισμούς

$$\min_w \frac{\|w\|^2}{2} + C \left(\sum_{i=1}^N \xi_i \right)^k$$

με τους περιορισμούς ότι

$$\begin{cases} y_i(w \cdot x_i + b) \geq 1 - \xi_i, & i = 1, 2, \dots, N \\ \xi_i > 0, & i = 1, 2, \dots, N \end{cases}$$

Με την ίδια λογική όπως πριν, το πρόβλημα βελτιστοποίησης μπορεί να λυθεί χρησιμοποιώντας την καθιερωμένη μέθοδο των πολλαπλασιαστών Lagrange.

- *Μη Γραμμικές SVMs*

Οι διατυπώσεις των SVMs που περιγράφηκαν στις δύο προηγούμενες περιπτώσεις κατασκευάζουν ένα γραμμικό σύνορο απόφασης για να διαχωρίσουν τα παραδείγματα εκπαίδευσης στις αντίστοιχες κλάσεις τους. Σε αυτή την περίπτωση θα παρουσιαστεί μία μεθοδολογία για την εφαρμογή των SVMs σε σύνολα δεδομένων που έχουν μη γραμμικά σύνορα απόφασης. Το τέχνασμα εδώ είναι να μετασχηματίσουμε τα δεδομένα από τον αρχικό χώρο συντεταγμένων x σε ένα νέο χώρο $\Phi(x)$ έτσι ώστε ένα γραμμικό σύνορο απόφασης να μπορεί να χρησιμοποιηθεί για να διαχωρίσει τα παραδείγματα στο μετασχηματισμένο χώρο. Μετά την πραγματοποίηση του μετασχηματισμού, μπορούμε να εφαρμόσουμε τη μεθοδολογία που παρουσιάστηκε στις



προηγούμενες περιπτώσεις για να βρούμε ένα γραμμικό σύνορο απόφασης στο μετασχηματισμένο χώρο.

Αν και η προσέγγιση του μετασχηματισμού των γνωρισμάτων φαίνεται υποσχόμενη, εγείρει διάφορα θέματα υλοποίησης. Πρώτα, δεν είναι σαφές τι τύπος συνάρτησης απεικόνισης θα πρέπει να χρησιμοποιηθεί για να εξασφαλιστεί ότι ένα γραμμικό σύνορο απόφασης μπορεί να κατασκευαστεί στο μετασχηματισμένο χώρο. Μία δυνατότητα είναι να μετασχηματίσουμε τα δεδομένα σε ένα χώρο άπειρης διάστασης, αλλά ένα τέτοιος χώρος τεράστιας διάστασης ίσως να μην είναι τόσο εύκολο να δουλέψουμε. Δεύτερο, ακόμα κι αν η κατάλληλη συνάρτηση απεικόνισης είναι γνωστή, η επίλυση του προβλήματος βελτιστοποίησης με περιορισμούς στον υψηλής διάστασης χώρο χαρακτηριστικών είναι ένα υπολογιστικά ακριβό έργο.

Για να διευκρινίσουμε αυτά τα ζητήματα και να εξετάσουμε τρόπους με τους οποίους μπορούν να διευθετηθούν, ας υποθέσουμε ότι υπάρχει μία κατάλληλη συνάρτηση $\Phi(x)$ για να μετασχηματίσουμε ένα δοσμένο σύνολο δεδομένων. Μετά το μετασχηματισμό, πρέπει να κατασκευάσουμε ένα γραμμικό σύνορο απόφασης που θα διαχωρίζει τα παραδείγματα στις αντίστοιχες κλάσεις. Το γραμμικό σύνορο απόφασης στο μετασχηματισμένο χώρο έχει την ακόλουθη μορφή

$$w \cdot \Phi(x) + b = 0$$

Σημείωση 2.3 (Μη Γραμμική SVM). Η διαδικασία μάθησης της SVM μπορεί να διατυπωθεί ως το ακόλουθο πρόβλημα βελτιστοποίησης με περιορισμούς

$$\min_w \frac{\|w\|^2}{2}$$

με τον περιορισμό ότι $y_i(w \cdot \Phi(x_i) + b) \geq 1, \quad i = 1, 2, \dots, N$

Παρατηρούμε την ομοιότητα μεταξύ της διαδικασίας μάθησης μίας μη γραμμικής SVM με αυτή μίας γραμμικής SVM. Η βασική διαφορά είναι ότι, αντί να χρησιμοποιηθούν τα αρχικά γνωρίσματα x , η διαδικασία μάθησης πραγματοποιείται στα μετασχηματισμένα γνωρίσματα $\Phi(x)$.

Παρομοίως με τις δύο προηγούμενες περιπτώσεις, το πρόβλημα βελτιστοποίησης μπορεί να λυθεί χρησιμοποιώντας την καθιερωμένη μέθοδο των πολλαπλασιαστών Lagrange.

Οι SVMs έχουν πολλές επιθυμητές ιδιότητες που τις κάνουν μία από τις πιο ευρέως χρησιμοποιούμενες αλγοριθμικές τεχνικές ταξινόμησης. Ακολουθούν αναφορικά κάποια γενικά χαρακτηριστικά τους.

- Το πρόβλημα μάθησης μίας SVM μπορεί να διατυπωθεί ως ένα κυρτό πρόβλημα βελτιστοποίησης, στο οποίο είναι διαθέσιμοι αποδοτικοί αλγόριθμοι που βρίσκουν το ολικό ελάχιστο της αντικειμενικής συνάρτησης. Άλλες μέθοδοι ταξινόμησης, όπως οι ταξινομητές βασισμένοι σε κανόνες και τα τεχνητά νευρωνικά δίκτυα, χρησιμοποιούν μία "άπληστη" στρατηγική για να ψάξουν το χώρο υπόθεσης. Τέτοιες μέθοδοι τείνουν να βρίσκουν μόνο τοπικά βέλτιστες λύσεις.



- Η SVM εκτελεί έλεγχο ικανότητας μεγιστοποιώντας το περιθώριο του συνόρου απόφασης. Παρόλα αυτά ο χρήστης πρέπει ακόμα να προμηθεύσει άλλες παραμέτρους όπως τον τύπο της συνάρτησης πυρήνα που θα χρησιμοποιηθεί και τη συνάρτηση κόστους C για την εισαγωγή κάθε μεταβλητής χαλάρωσης.
- Η SVM μπορεί να εφαρμοστεί σε κατηγορικά δεδομένα εισάγοντας ψεύτικες μεταβλητές για κάθε κατηγορική τιμή γνωρίσματος που υπάρχει στα δεδομένα. Για παράδειγμα, αν η το γνώρισμα 'συζυγική κατάσταση' έχει τρεις τιμές {Ελεύθερος, Παντρεμένος, Χωρισμένος}, μπορούμε να εισάγουμε μία δυαδική μεταβλητή για κάθε μία από τις τιμές του γνωρίσματος 'συζυγική κατάσταση'.
- Η διατύπωση της SVM που παρουσιάστηκε είναι για προβλήματα δύο κλάσεων. Ωστόσο, υπάρχουν μέθοδοι που επεκτείνουν τις SVMs σε προβλήματα πολλών κλάσεων. Αυτές εφαρμόζονται και στα πειράματα της παρούσας εργασίας που χρησιμοποιούν τεχνικές SVM.



ΚΕΦΑΛΑΙΟ 3

ΚΑΤΑΤΜΗΣΗ ΕΙΚΟΝΑΣ ΜΕ ΜΕΘΟΔΟΥΣ ΜΗ ΕΠΙΒΛΕΠΟΜΕΝΗΣ ΜΑΘΗΣΗΣ

3.1 Εισαγωγή

3.2 Βάση του Berkeley

3.3 Αλγόριθμος

3.4 Πειραματικά αποτελέσματα

3.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται η μη επιβλεπόμενη μάθηση και η εφαρμογή της για την ανίχνευση ομοιογενών περιοχών σε εικόνες. Στην παράγραφο 3.2 περιγράφεται η βάση του Berkeley που χρησιμοποιήθηκε στα πειράματα της μη επιβλεπόμενης μάθησης. Στην παράγραφο 3.3 περιγράφεται αναλυτικά ο βασικός αλγόριθμος μη επιβλεπόμενης μάθησης. Τέλος, στην παράγραφο 3.4 παρουσιάζονται τα πειραματικά αποτελέσματα της μεθόδου που χρησιμοποιήθηκε.

3.2 Βάση του Berkeley

Η βάση του Berkeley [11] περιέχει ένα μεγάλο πλήθος εικόνων που μπορούν να χρησιμοποιηθούν στην περιοχή της κατάτμησης εικόνας. Οι εικόνες που περιλαμβάνονται έχουν επιπλέον καταταμηθεί με χειρόγραφες ετικέτες από 30 διαφορετικά άτομα. Το τμήμα της βάσης του Berkeley, που διατίθεται δημοσίως για ερευνητικούς σκοπούς, αποτελείται από 300 εικόνες που υπάρχουν τόσο σε ασπρόμαυρη όσο και σε έγχρωμη μορφή.



Για κάθε μία από τις 300 έγχρωμες εικόνες υπάρχουν χειρόγραφες εικόνες με ετικέτες, οι οποίες δημιουργούνται ξεχωριστά και ανεξάρτητα συνήθως από 5 έως 8 άτομα. Τα συγκεκριμένα άτομα προσπάθησαν να χωρίσουν την εικόνα σε κάποιες ομοιογενείς περιοχές που αναπαριστούν κατά την άποψή τους διαφορετικά αντικείμενα και έννοιες στην εικόνα. Αυτές οι εικόνες ετικετών χρησιμοποιούνται μετέπειτα για τον υπολογισμό δεικτών που αφορούν την ποιότητα της κατάτμησης εικόνας που πραγματοποιήσαμε. Επιπλέον, για κάθε μία από τις εικόνες διατίθεται κατάτμηση σε υπερεικονοστοιχεία, η οποία πραγματοποιείται όπως περιγράφηκε στην παράγραφο (2.5.1). Αυτά τα υπερεικονοστοιχεία αποτελούν τη βάση των αλγορίθμων που χρησιμοποιήθηκαν.

Στην παρούσα εργασία πραγματοποιήθηκαν πειράματα με το σύνολο των 300 εικόνων. Επιπλέον, οι δείκτες που αφορούν την ποιότητα της κατάτμησης εικόνας υπολογίστηκαν συνολικά στις 300 εικόνες. Στην παράγραφο (3.4) θα αναφερθούν επιπλέον αναλυτικά κάποιες διαπιστώσεις σχετικά με τα αριθμητικά αποτελέσματα των δεικτών που προέκυψαν.

Στο σχήμα (3.1) φαίνονται κάποια παραδείγματα εικόνων της βάσης του Berkeley και η δοσμένη κατάτμηση τους σε υπερεικονοστοιχεία.

3.3 Αλγόριθμος

Στόχος του αλγορίθμου μη επιβλεπόμενης μάθησης σε επίπεδο εικόνας είναι η ομαδοποίηση των υπερεικονοστοιχείων της σε ευρύτερα ομοιογενείς περιοχές. Η ομοιογένεια κάθε ευρύτερης περιοχής αφορά χαρακτηριστικά και ιδιότητες, όπως απόσταση, χρώμα, φωτεινότητα, υφή κ.λ.π. Επιπλέον, θα πρέπει διαφορετικές περιοχές να είναι διακριτές μεταξύ τους όσον αφορά τα αντίστοιχά τους χαρακτηριστικά. Ιδεατά επιθυμούμε οι περιοχές που θα προκύψουν μετά τη συνένωση των υπερεικονοστοιχείων να αντιστοιχούν σε αντικείμενα και έννοιες της εικόνας. Δηλαδή, η εικόνα να είναι κατετμημένη σε περιοχές που μπορούν να χαρακτηριστούν ανεξάρτητες μεταξύ τους και οι οποίες να αντιπροσωπεύουν ξεχωριστά οντότητες του πραγματικού κόσμου. Συμπερασματικά, επιδιώκουμε να προσεγγίσουν όσο το δυνατόν καλύτερα κάθε πιθανή χειρόγραφη κατάτμηση της εικόνας από κάποιο τυχαίο άτομο, το οποίο θα αντιστοιχούσε κάθε οπτικά διακριτή περιοχή της εικόνας με μία διαφορετική ετικέτα.

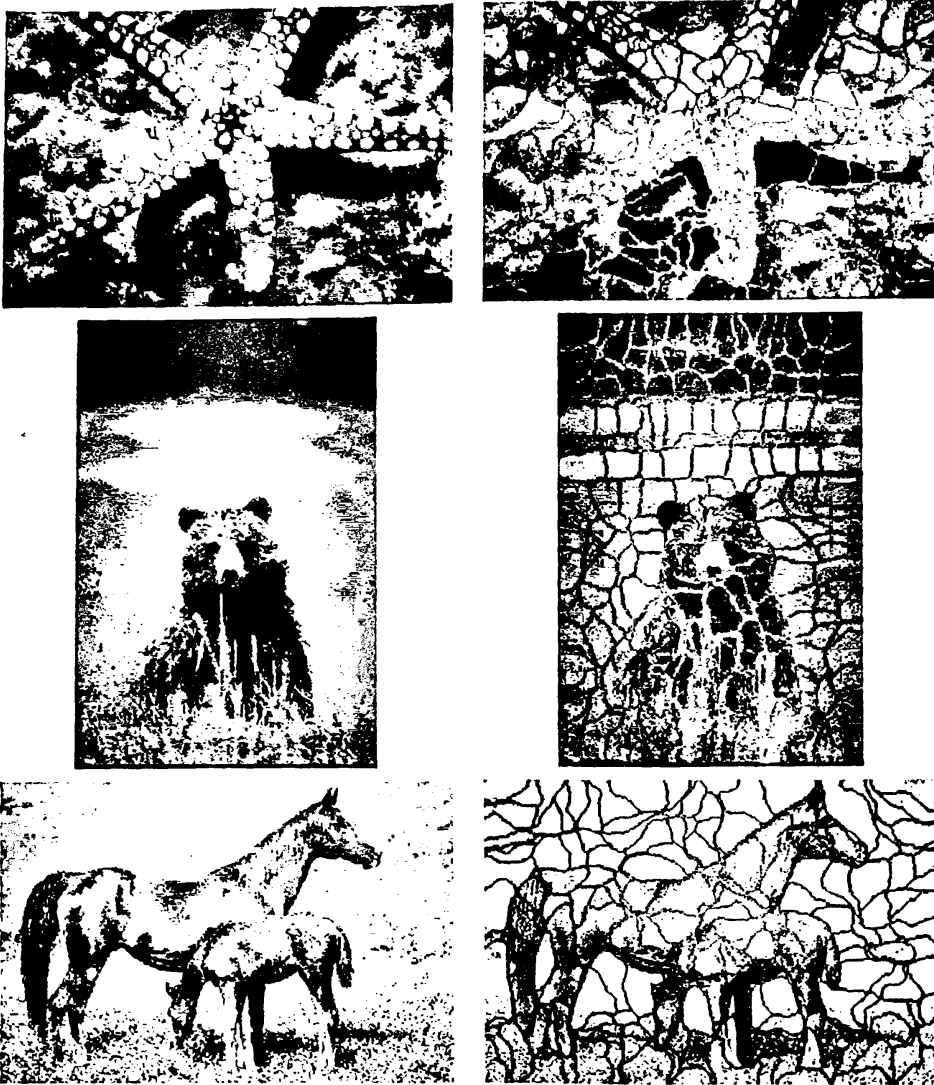
Παρακάτω ακολουθεί συνοπτικά η περιγραφή του αλγορίθμου (5) στο πλαίσιο της ομαδοποίησης των υπερεικονοστοιχείων μιας εικόνας σε ευρύτερες περιοχές.

Εναλλακτικά, μπορούμε να τροποποιήσουμε το βήμα 4 του αλγορίθμου (5), όπου αντί για την απόκλιση KL, χρησιμοποιούμε την απόσταση κοντινότερου γείτονα

$$d_{ij}(\mu_i, \mu_j) = (\mu_i - \mu_j)^T (\mu_i - \mu_j)$$

Οι διαφορετικές παραλλαγές του αλγορίθμου (5) παρουσιάζονται με εικόνες και πίνακες αξιολόγησης κατάτμησης στα πειραματικά αποτελέσματα (3.4).





Σχήμα 3.1: Εικόνες της βάσης του Berkeley και τα δοθέντα υπερεικονοστοιχεία

3.4 Πειραματικά αποτελέσματα

Στον αλγόριθμο μη επιβλεπόμενης μάθησης (5) μελετήθηκαν πειραματικά περαιτέρω κάτω παράμετροι προκειμένου να βελτιωθεί κατά το δυνατόν η ποιότητα της και τόσο οπτικά όσο και βάσει των δεικτών

1. αριθμός ομάδων M .
2. σταθερά c στην ομοιότητα.
3. σταθερά a στη γειτνίαση των υπερεικονοστοιχείων.

Ακολουθούν κάποιες παρατηρήσεις σχετικά με τις παραμέτρους και την επίδραση στο τελικό αποτέλεσμα της κατάτμησης

- ο αριθμός ομάδων M καθορίστηκε να παίρνει τις εξής τιμές $\{3,5,8,10,15,20\}$.
 νες της βάσης του Berkeley κατατμήθηκαν με αποτέλεσμα να προκύψουν M

Αλγόριθμος 5 Αλγόριθμος μη επιβλεπόμενης μάθησης

- 1 Δοθείσης εικόνας I , χρησιμοποιούμε τα δοσμένα υπερεικονοστοιχεία της. Αυτά είχαν προκύψει εφαρμόζοντας τον αλγόριθμο Neut.
- 2 Εξάγουμε τα MRF χαρακτηριστικά της εικόνας, εφαρμόζουμε PCA για μείωση διάστασης και τα κανονικοποιούμε.
- 3 Μοντελοποιούμε κάθε υπερεικονοστοιχείο με κανονική κατανομή $N(\mu, \Sigma)$ με μέση τιμή μ το μέσο όρο των τιμών των MRF χαρακτηριστικών των εικονοστοιχείων του και πίνακα συμμεταβλητότητας Σ τον πίνακα συνδιακύμανσης που αντιστοιχεί σε αυτές τις τιμές.
- 4 Κατασκευάζουμε τον πίνακα ομοιότητας W των υπερεικονοστοιχείων i, j

$$w_{ij} = \exp\left(-\frac{d_{ij}}{c}\right)$$

όπου c σταθερά και d η απόκλιση KL

$$d_{ij}(N(\mu_i, \Sigma_i), N(\mu_j, \Sigma_j)) = (\mu_i - \mu_j)^T(\mu_i - \mu_j) + \text{Tr}\left(\Sigma_i + \Sigma_j - 2(\Sigma_i \Sigma_j)^{\frac{1}{2}}\right)$$

- 5 Προσαρμόζουμε τον πίνακα ομοιότητας W ώστε να εμπεριέχει πληροφορία γειτνίασης των υπερεικονοστοιχείων

$$w_{ij} = \begin{cases} a * w_{ij} & \text{αν } i, j \text{ μη γειτονικά υπερεικονοστοιχεία} \\ w_{ij} & \text{αν } i, j \text{ γειτονικά υπερεικονοστοιχεία} \end{cases}$$

όπου $a \in [0, 1]$ σταθερά

- 6 Υπολόγισε τις μικρότερες $M + 1$ ιδιοτιμές $\lambda_0, \lambda_1, \dots, \lambda_M$ του κανονικοποιημένου Λαπλασιανού πίνακα $\bar{L} \equiv D^{-1/2} L D^{-1/2}$ (2.5.1) και τοποθέτησε τα αντίστοιχα ιδιοδιανύσματα, εκτός του πρώτου, u^1, \dots, u^M σε ένα πίνακα U ως στήλες, όπου M καθορίζεται από το χρήστη.
- 7 Θεωρώντας κάθε γραμμή i του πίνακα U ως διάνυσμα αναπαράστασης του υπερεικονοστοιχείου i , εφαρμόζουμε τον αλγόριθμο K -μέσων (k -means) στις γραμμές του U και καθορίζουμε τον αριθμό ομάδων ίσο με M .
- 8 Κατατάσσουμε τα υπερεικονοστοιχεία σε ευρύτερες περιοχές σύμφωνα με την ομαδοποίηση που παράγεται από την εφαρμογή του αλγορίθμου K -μέσων.

ανά εικόνα. Για τις τιμές $\{3, 5, 8\}$ προέκυψαν καλές οπτικές κατατιμήσεις, καθώς συνήθως κάθε εικόνα θα μπορούσε να χαρακτηριστεί χειρόγραφα σε 3 έως 8 ευδιάκριτες περιοχές. Για τις τιμές $\{10, 15, 20\}$ παρατηρούμε ότι οι κατατιμήσεις περιέχουν υπο-



περιοχές ομογενών ευρύτερων περιοχών. Αυτό έχει ως αποτέλεσμα οι υποπεριοχές να αποτελούν τμήμα κάποιου αντικειμένου ή έννοιας της εικόνας και, επομένως, να μην μπορούν να χαρακτηριστούν οπτικά. Ωστόσο, παραμένουν ομοιογενείς και δεν επηρεάζουν σημαντικά προς το χειρότερο τους δείκτες αξιολόγησης της κατάτμησης.

- η σταθερά c στην ομοιότητα

$$w_{ij} = \exp\left(-\frac{d_{ij}}{c}\right)$$

παίρνει τις εξής τιμές $\{10,100,1000\}$. Επιπλέον, μελετήθηκε και η ομοιότητα

$$w_{ij} = \exp\left(-\frac{\sqrt{d_{ij}}}{c'}\right)$$

όπου η σταθερά c' παίρνει παρομοίως τις τιμές $\{10,100,1000\}$.

Διαπιστώθηκε ότι το είδος της ομοιότητας επηρεάζει σημαντικά την κατάτμηση. Πειραματικά προέκυψε ότι η καλύτερη ομοιότητα είναι η

$$w_{ij} = \exp\left(-\frac{\sqrt{d_{ij}}}{10}\right)$$

Βέβαια, στην περίπτωση που χρησιμοποιείται η απόσταση κοντινότερου γείτονα (2.26) (όπου δεν υπάρχει ο πίνακας συμμεταβλητότητας Σ), το είδος ομοιότητας και η σταθερά της παίζει λιγότερο ρόλο. Αυτό συμβαίνει καθώς προηγουμένως ο πίνακας Σ περιείχε μεγάλες τιμές σε μη διαγώνια στοιχεία. Επομένως, υπήρχε μεγάλη συμμεταβλητότητα διαφορετικών υπερεικονοστοιχείων της εικόνας και, επομένως, η ομοιότητα και η σταθερά της συνέβαλαν καθοριστικά στον προσδιορισμό της μεταξύ τους σχέσης.

- η σταθερά a στη γειτνίαση των υπερεικονοστοιχείων καθορίστηκε να παίρνει τις εξής τιμές $\{0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0\}$. Παρατηρήθηκε ότι μία "μέση" λύση, δηλαδή οι τιμές $\{0.4,0.5,0.6\}$, έδωσαν γενικά τα καλύτερα αποτελέσματα. Ωστόσο, η επίδραση της a δεν ήταν τόσο σημαντική όσο προηγουμένως η επιρροή της ομοιότητας και της σταθεράς της. Επιπλέον, η ιδανική τιμή της a διαφέρει αρκετά ανά εικόνα με τη γενικότερη τάση να επικρατούν τιμές μεγαλύτερες του 0.5. Αυτό σημαίνει ότι η ομοιότητα των μη γειτονικών υπερεικονοστοιχείων δε θα πρέπει να μειώνεται κάτω του μισού, αλλά αντίθετα να διατηρείται σε ένα σημαντικό ποσοστό. Η διαπίστωση αυτή τεκμηριώνεται και λογικά, καθώς μπορεί να υπάρχουν υπερεικονοστοιχεία που να απέχουν αρκετά χωρικά, ωστόσο να ανήκουν στην ίδια ευρύτερη ομοιογενή περιοχή και, επομένως, θα πρέπει να ομαδοποιηθούν μεταξύ τους.

Όσον αφορά την αξιολόγηση της παραγόμενης κατάτμησης, χρησιμοποιούνται οι εξής δείκτες

- Ο δείκτης Boundary Displacement Error (BDE) [4] μετρά το μέσο σφάλμα μετατόπισης των οριακών εικονοστοιχείων μεταξύ δύο κατετμημένων εικόνων. Συγκεκριμένα,



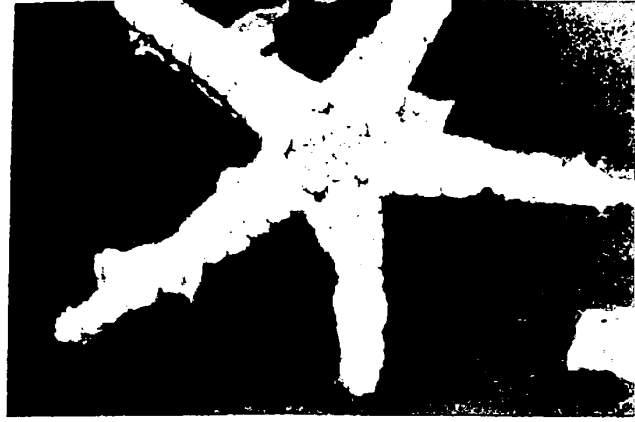
καθορίζει το σφάλμα ενός οριακού εικονοστοιχείου ως την απόσταση μεταξύ αυτού του εικονοστοιχείου και του κοντινότερου οριακού εικονοστοιχείου στην άλλη εικόνα.

- Ο δείκτης Probabilistic Rand Index (PRI) [16] μετράει το ποσοστό των ζευγών εικονοστοιχείων των οποίων οι ετικέτες είναι σταθερές μεταξύ της υπολογισμένης κατάτμησης και της χειρόγραφης κατάτμησης με ετικέτες, παίρνοντας το μέσο όρο των χειρόγραφων κατατμήσεων που επιχείρησαν διαφορετικά άτομα προκειμένου να ληφθεί υπόψη η κλίμακα διακύμανσης της ανθρώπινης αντίληψης.
- Ο δείκτης Variation Of Information (VOI) [13] καθορίζει την απόσταση μεταξύ δύο κατατμήσεων ως τη μέση υπό συνθήκη εντροπία της μίας κατάτμησης δοθείσης της άλλης και, επομένως, μετράει περίπου το ποσοστό της τυχαιότητας στη μία κατάτμηση το οποίο δεν μπορεί να εξηγηθεί από την άλλη.
- Ο δείκτης Global Consistency Error (GCE) [10] μετράει το βαθμό στον οποίο μία κατάτμηση μπορεί να θεωρηθεί ως μία επαναδιάταξη της άλλης. Οι κατατμήσεις που σχετίζονται με αυτόν τον τρόπο θεωρούνται ότι είναι σταθερές, καθώς θα μπορούσαν να αναπαριστούν την ίδια φυσική εικόνα κατατμημένη σε διαφορετικές κλίμακες.

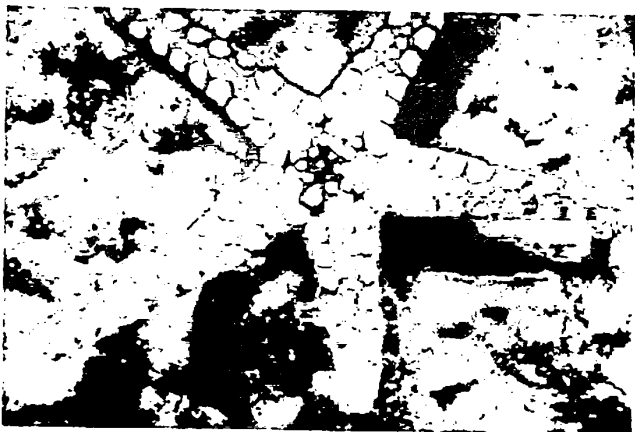
Παρακάτω παρατίθενται διάφορες εικόνες της βάσης του Berkeley και οι κατατμήσεις τους καθώς και πίνακες με τους δείκτες αξιολόγησης της ποιότητας των κατατμήσεων.

Συγκεκριμένα, παρουσιάζονται εικόνες μόνο με χρήση Ncut, επειδή ο αλγόριθμος Ncut (2.5.1) υπερσχύει σε απόδοση έναντι του αλγορίθμου με χρήση του πίνακα ομοιότητας (2.5.2). Οι εικόνες αφορούν την εφαρμογή Ncut για φασματική ομαδοποίηση σε επίπεδο υπερεικονοστοιχείου, όπου εφαρμόζεται ο αλγόριθμος (5) της ενότητας (3.3). Παρουσιάζονται εικόνες τόσο με χρήση της απόκλισης KL όσο και με χρήση της απόστασης κοντινότερου γείτονα. Επιπλέον, παρουσιάζονται δείκτες αξιολόγησης της ποιότητας των κατατμήσεων με χρήση Ncut (2.5.1) τόσο με απόκλιση KL όσο και με απόσταση κοντινότερου γείτονα. Επιπρόσθετα, παρατίθενται δείκτες αξιολόγησης με χρήση κατάτμησης με χρήση του πίνακα ομοιότητας (2.5.2) και απόκλισης KL, προκειμένου να δειχτεί ότι ο Ncut υπερτερεί.



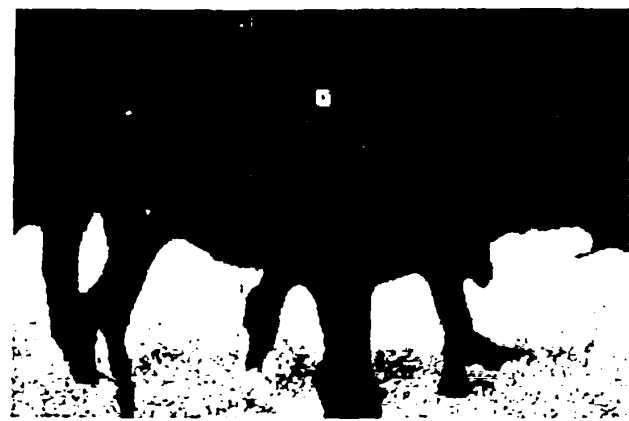
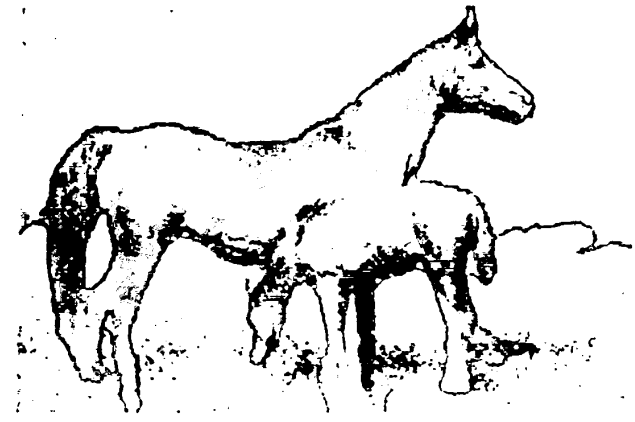


2: Η κατάτμηση της εικόνας "Αστερίας" σε 3 ομάδες με χρήση της απόστασης από την εικόνα $a=0.3$.

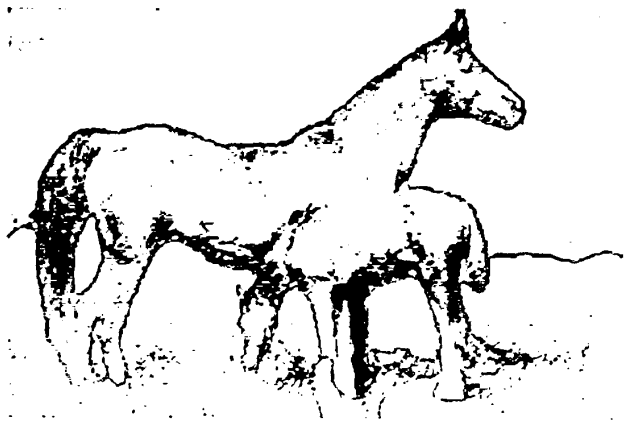


3: Η κατάτμηση της εικόνας "Αστερίας" σε 3 ομάδες με χρήση της απόστασης από την εικόνα $a=0.3$.





4: Η κατάτμηση της εικόνας "Άλογα" σε 3 ομάδες με χρήση της απόκλισης. Παράμετρος $a=0.9$.



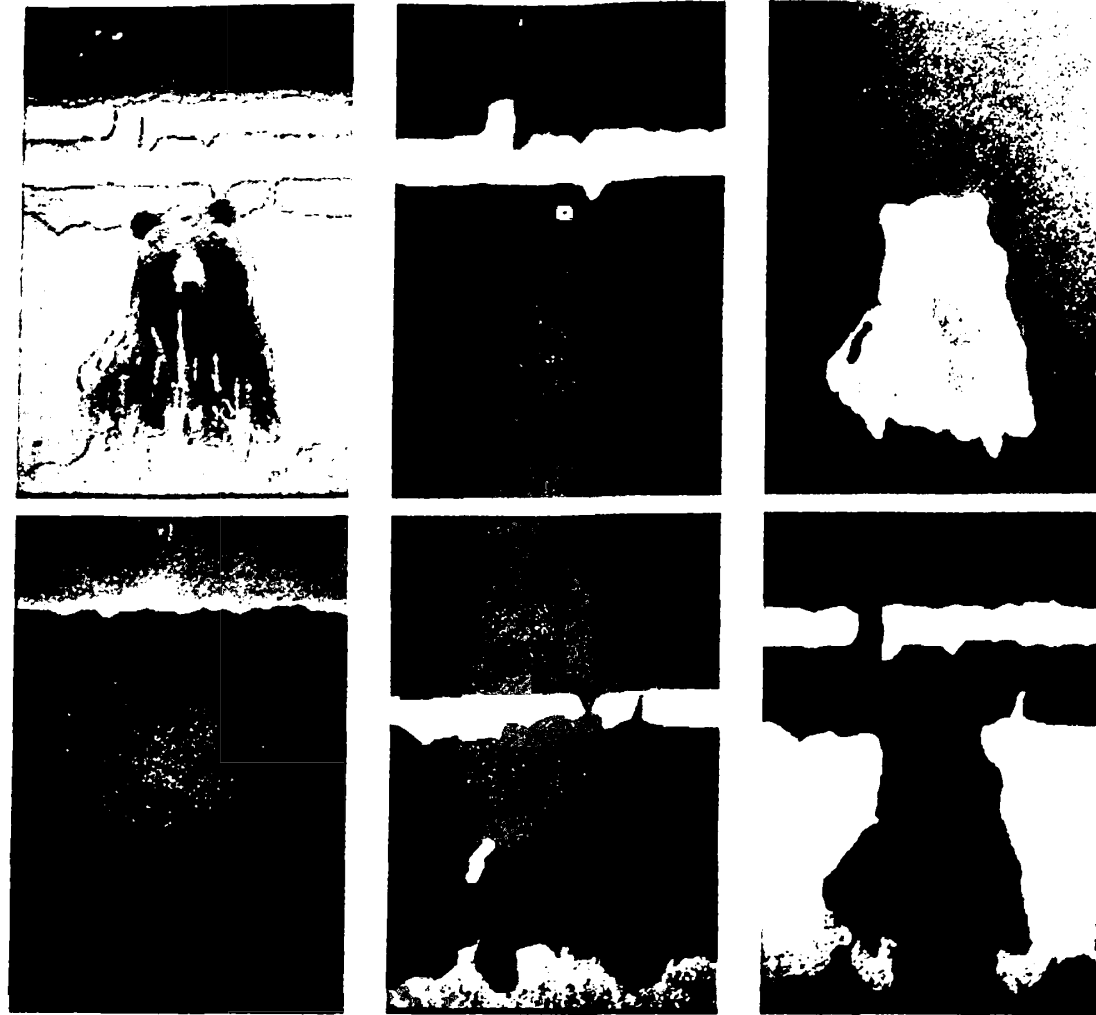
5: Η κατάτμηση της εικόνας "Άλογα" σε 3 ομάδες με χρήση της απόστασης. Βέλτιστη παράμετρος $a=0.9$.



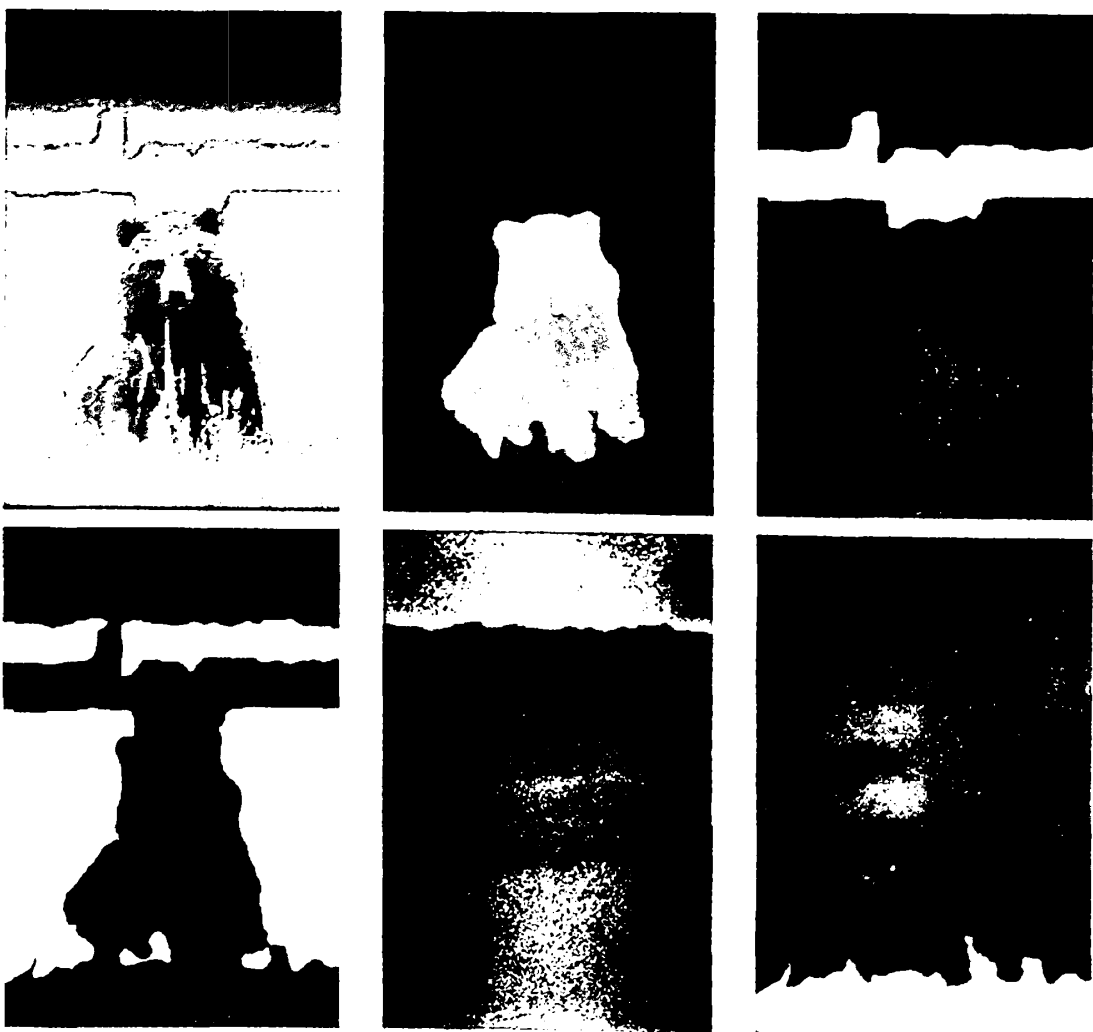


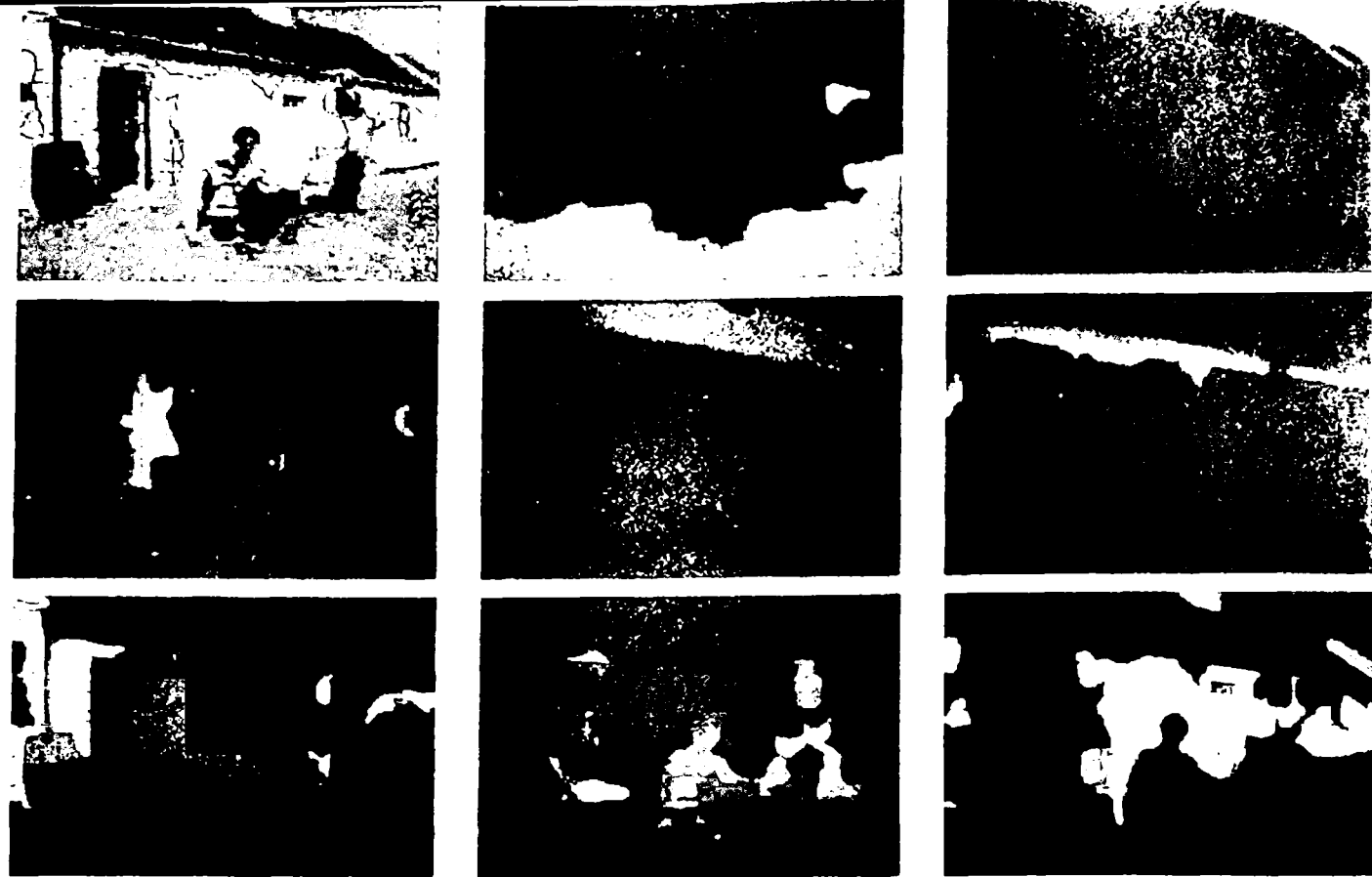
Η κατάτμηση της εικόνας "Ελέφαντας" σε 5 ομάδες με χρήση της κλίμακας $a=0.7$.



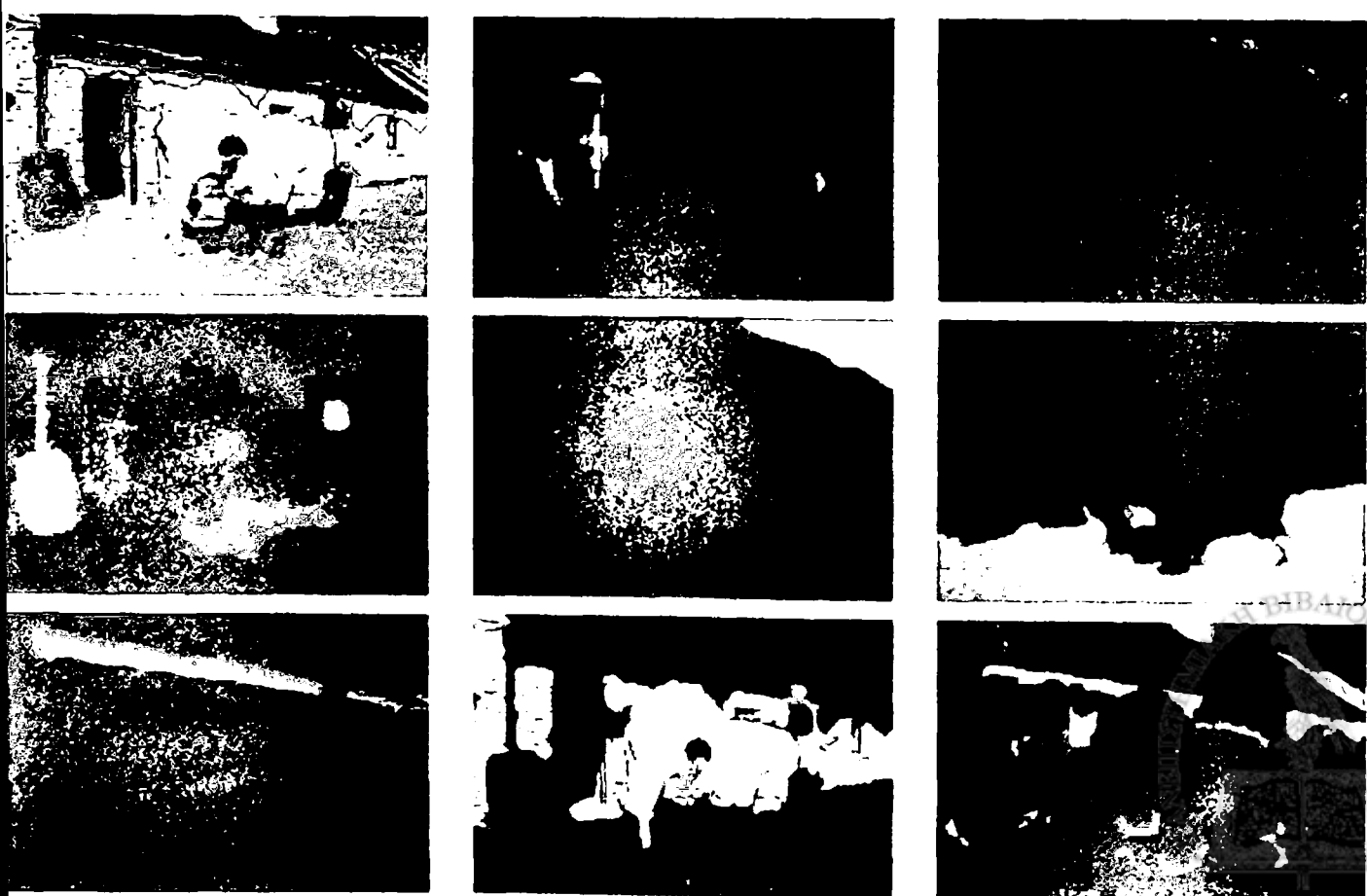


Η κατάτμηση της εικόνας "Αρχούδα" σε 5 ομάδες με χρήση της $\alpha=0.7$.





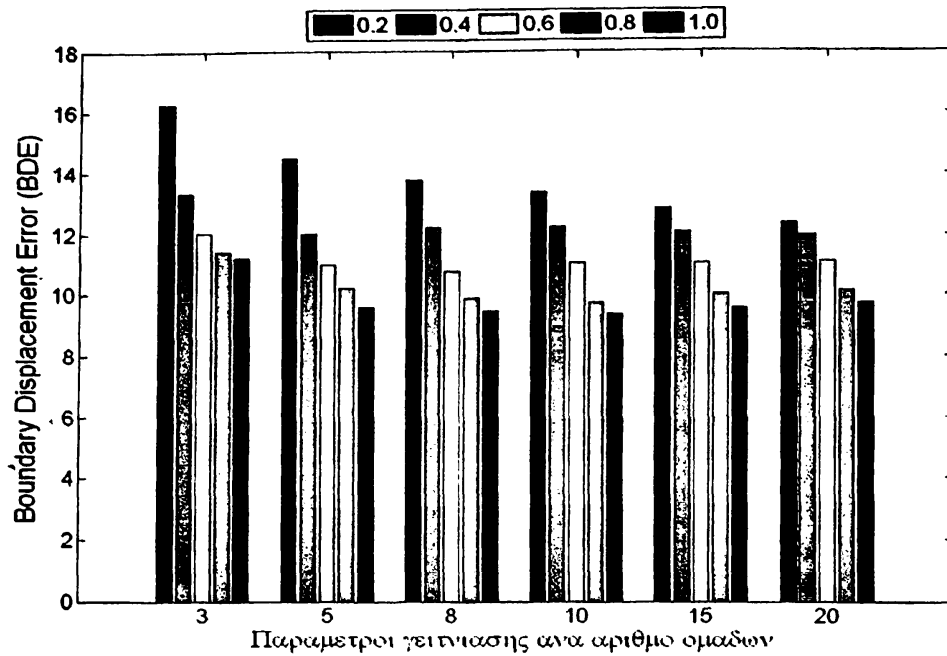
10: Η κατάτμηση της εικόνας "Αγρόκτημα" σε 8 ομάδες με χρήση της απόκλιση παράμετρος $a=0.7$.



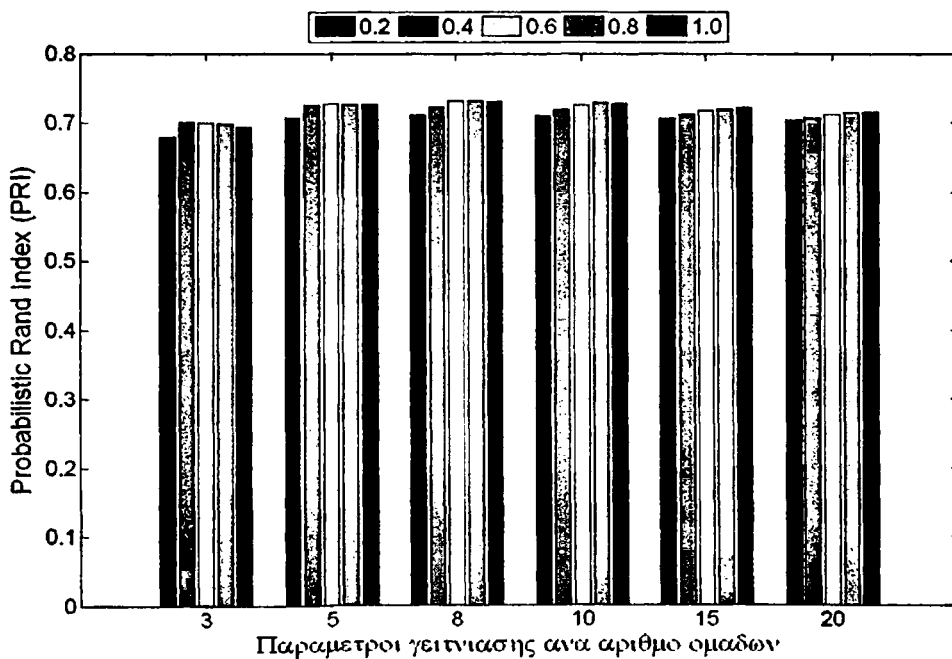
Πίνακας 3.1: Αξιολόγηση των κατατμήσεων με τη μέθοδο Neut και απόκλιση KL. Συμβολίζουμε με M τον αριθμό ομάδων και με a τη σταθερά γειτνίασης των υπερεικονοστοιχείων.

$M \backslash a$		0.2	0.4	0.6	0.8	1.0
		3	BDE	16.2539	13.3179	12.0470
	PRI	0.6795	0.7014	0.6992	0.6988	0.6948
	VOI	2.2976	2.1606	2.1782	2.2038	2.2343
	GCE	0.2519	0.2210	0.2241	0.2281	0.2333
5	BDE	14.4962	12.0084	10.9837	10.2256	9.5845
	PRI	0.7068	0.7254	0.7280	0.7275	0.7263
	VOI	2.5679	2.4049	2.3957	2.4577	2.5007
	GCE	0.3425	0.3104	0.3089	0.3195	0.3258
8	BDE	13.7723	12.2125	10.7659	9.8738	9.4446
	PRI	0.7115	0.7236	0.7314	0.7315	0.7308
	VOI	2.8542	2.6961	2.6569	2.7274	2.7960
	GCE	0.3575	0.3286	0.3216	0.3352	0.3490
10	BDE	13.3644	12.2317	11.0123	9.7319	9.3690
	PRI	0.7102	0.7191	0.7253	0.7300	0.7282
	VOI	3.0144	2.8680	2.8158	2.8551	2.9435
	GCE	0.3511	0.3245	0.3152	0.3250	0.3474
15	BDE	12.8295	12.0857	11.0380	10.0042	9.5614
	PRI	0.7067	0.7121	0.7182	0.7205	0.7220
	VOI	3.3127	3.1813	3.0821	3.1313	3.2226
	GCE	0.3247	0.2999	0.2840	0.2968	0.3278
20	BDE	12.3389	11.9644	11.1182	10.1385	9.7479
	PRI	0.7052	0.7078	0.7120	0.7152	0.7160
	VOI	3.4999	3.4149	3.3148	3.3114	3.4385
	GCE	0.2940	0.2790	0.2617	0.2668	0.3096



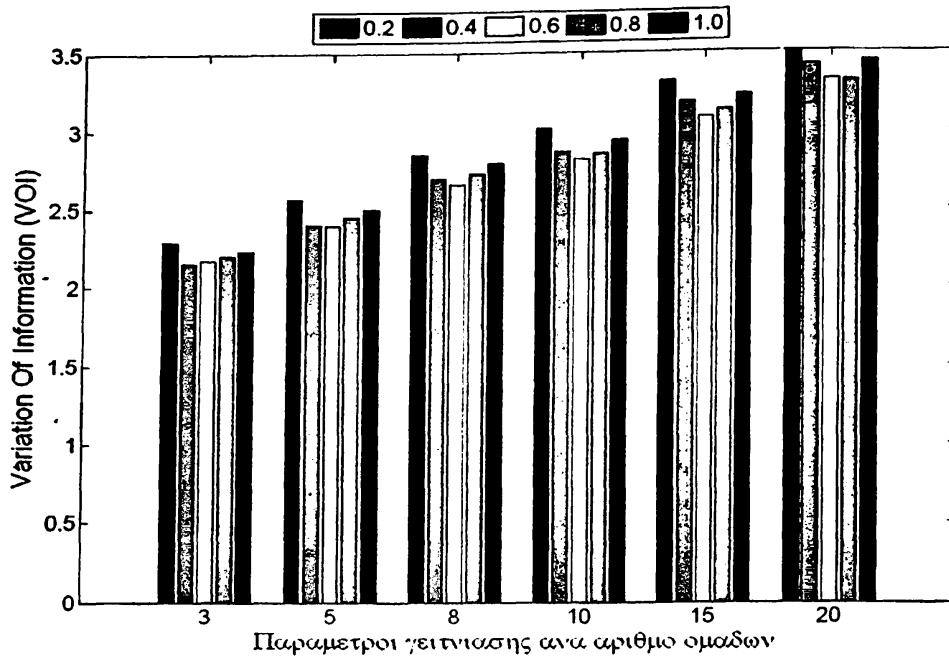


Σχήμα 3.12: Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόκλιση KL. Ο δείκτης BDE. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης α που φαίνονται στο πάνω μέρος του γραφήματος.

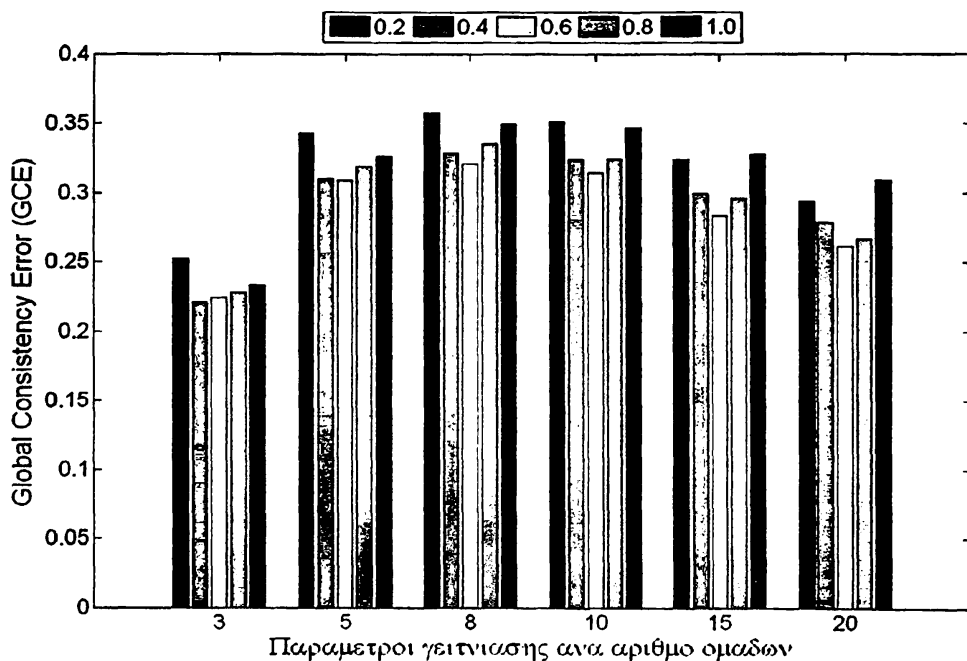


Σχήμα 3.13: Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόκλιση KL. Ο δείκτης PRI. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης α που φαίνονται στο πάνω μέρος του γραφήματος.





Σχήμα 3.14: Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόκλιση KL. Ο δείκτης VOI. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης α που φαίνονται στο πάνω μέρος του γραφήματος.



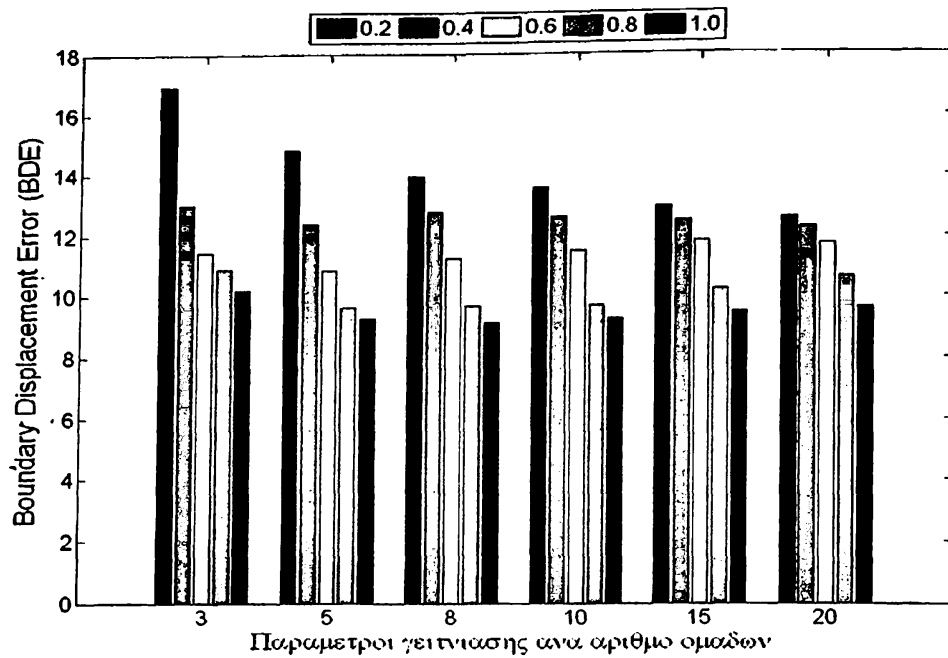
Σχήμα 3.15: Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόκλιση KL. Ο δείκτης GCE. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης α που φαίνονται στο πάνω μέρος του γραφήματος.



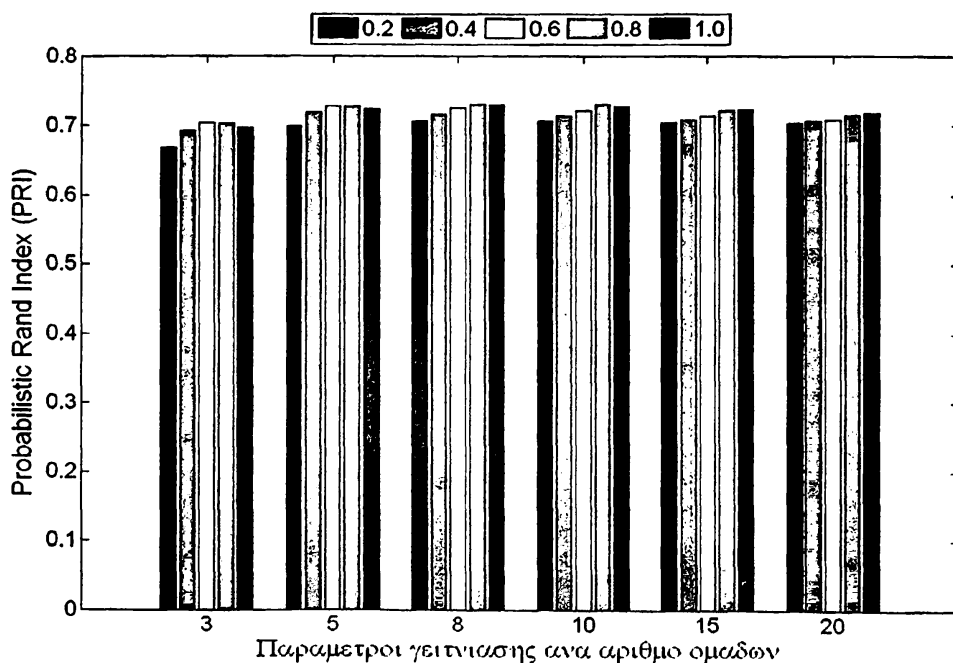
Πίνακας 3.2: Αξιολόγηση των κατατιμήσεων με τη μέθοδο Neut και απόσταση κοντινότερου γείτονα. Συμβολίζουμε με M τον αριθμό ομάδων και με a τη σταθερά γειτνίασης των υπεριοστοιχείων.

$M \backslash a$		0.2	0.4	0.6	0.8	1.0
		3	BDE	16.9251	13.0029	11.4512
	PRI	0.6670	0.6922	0.7031	0.7021	0.6965
	VOI	2.3855	2.2212	2.1708	2.1880	2.2407
	GCE	0.2706	0.2346	0.2213	0.2222	0.2315
5	BDE	14.8495	12.3817	10.8630	9.6465	9.2880
	PRI	0.6983	0.7197	0.7286	0.7277	0.7242
	VOI	2.6576	2.4679	2.4110	2.4502	2.4923
	GCE	0.3590	0.3200	0.3092	0.3134	0.3181
8	BDE	13.9334	12.7636	11.2473	9.6907	9.1439
	PRI	0.7058	0.7161	0.7255	0.7310	0.7298
	VOI	2.9511	2.8107	2.7165	2.7151	2.7731
	GCE	0.3743	0.3470	0.3310	0.3343	0.3489
10	BDE	13.5495	12.6031	11.4965	9.6876	9.2529
	PRI	0.7062	0.7142	0.7219	0.7303	0.7288
	VOI	3.0990	2.9582	2.8646	2.8310	2.8920
	GCE	0.3645	0.3390	0.3239	0.3227	0.3483
15	BDE	12.9090	12.4835	11.8188	10.2120	9.4703
	PRI	0.7051	0.7089	0.7137	0.7216	0.7229
	VOI	3.3602	3.2692	3.1587	3.0578	3.1424
	GCE	0.3324	0.3157	0.2960	0.2866	0.3322
20	BDE	12.5712	12.2371	11.7191	10.6279	9.6124
	PRI	0.7037	0.7058	0.7086	0.7151	0.7178
	VOI	3.5578	3.4850	3.4021	3.2456	3.3419
	GCE	0.3051	0.2914	0.2768	0.2568	0.3184



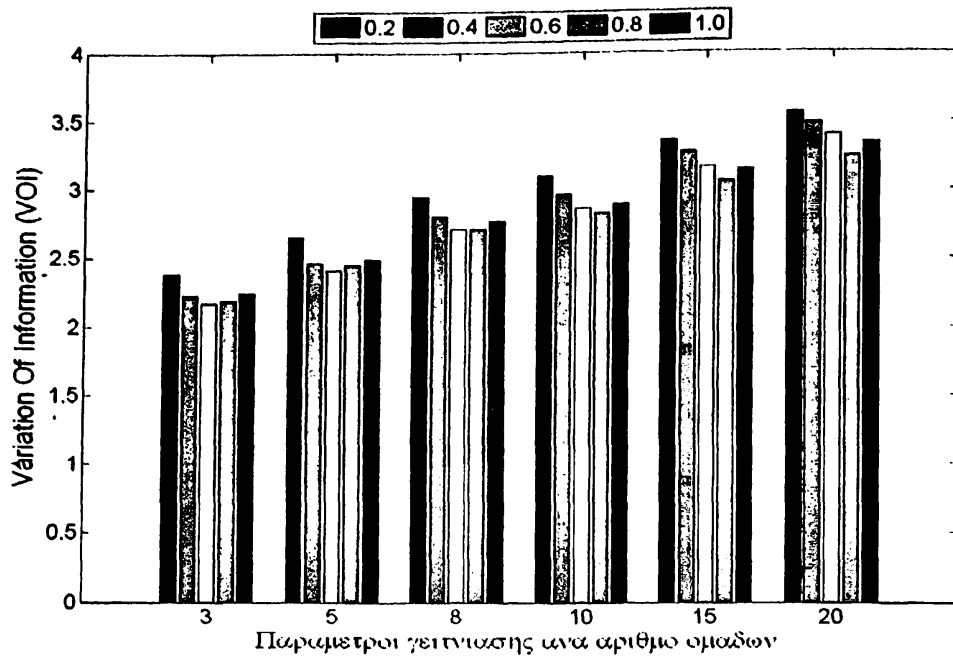


Σχήμα 3.16: Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόσταση κοντινότερου γείτονα. Ο δείκτης BDE. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης α που φαίνονται στο πάνω μέρος του γραφήματος.

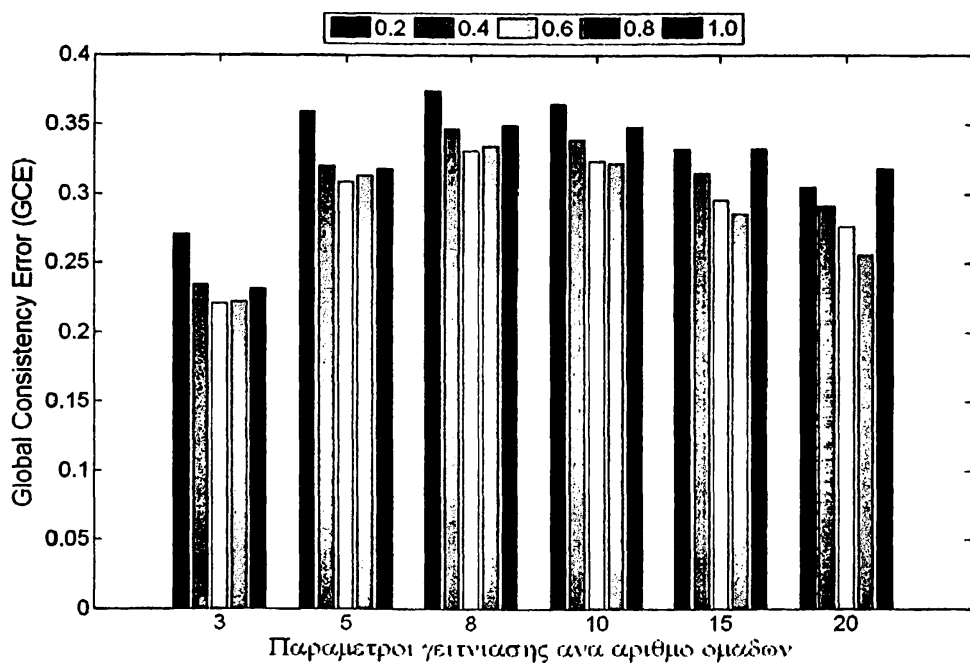


Σχήμα 3.17: Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόσταση κοντινότερου γείτονα. Ο δείκτης PRI. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης α που φαίνονται στο πάνω μέρος του γραφήματος.





Σχήμα 3.18: Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόσταση κοντινότερου γείτονα. Ο δείκτης VOI. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνιασης α που φαίνονται στο πάνω μέρος του γραφήματος.



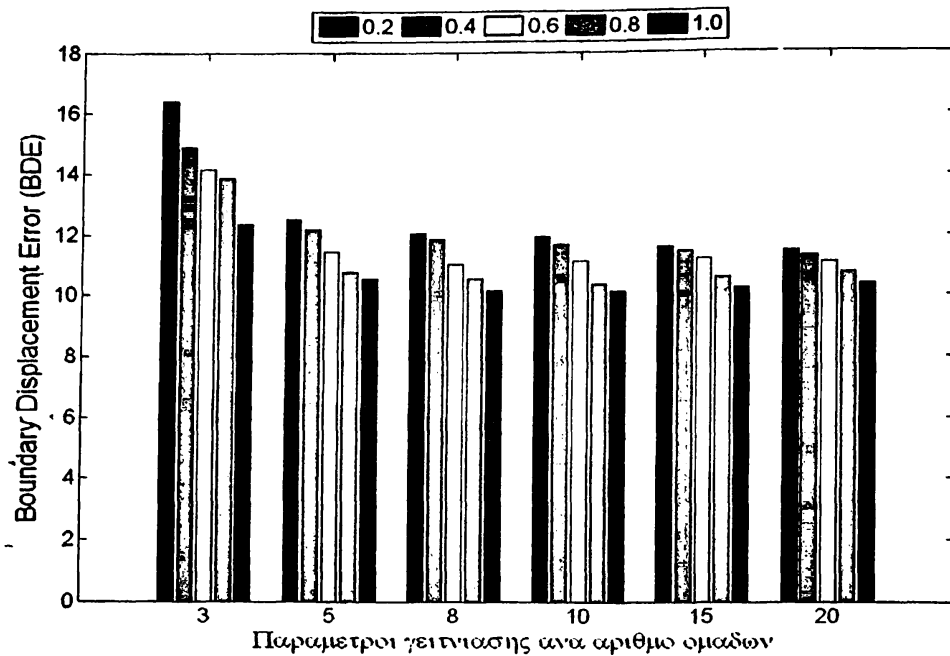
Σχήμα 3.19: Αξιολόγηση των κατατιμήσεων με τη μέθοδο Ncut και απόσταση κοντινότερου γείτονα. Ο δείκτης GCE. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνιασης α που φαίνονται στο πάνω μέρος του γραφήματος.



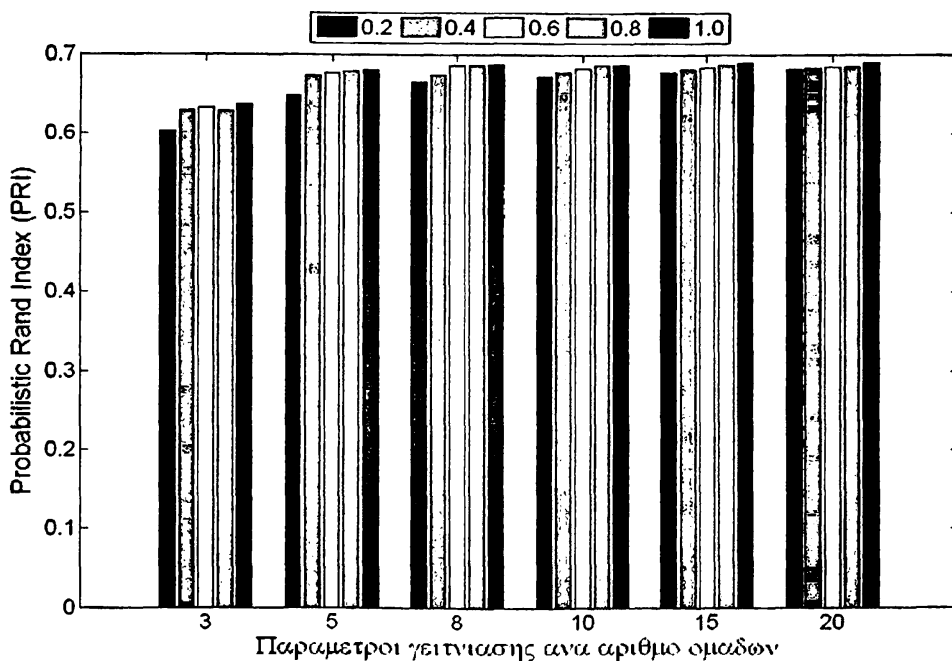
Πίνακας 3.3: Αξιολόγηση των κατατμήσεων με τη μέθοδο φασματικής ομαδοποίησης με χρήση του πίνακα ομοιότητας και απόκλιση KL. Συμβολίζουμε με M τον αριθμό ομάδων και με a τη σταθερά γειτνίασης των υπερεικονοστοιχείων.

$M \backslash a$		0.2	0.4	0.6	0.8	1.0
		3	BDE	16.3993	14.8864	14.1695
	PRI	0.6030	0.6297	0.6326	0.6283	0.6374
	VOI	2.7577	2.5752	2.5894	2.5816	2.6094
	GCE	0.3361	0.2997	0.3041	0.2986	0.3105
5	BDE	12.5074	12.2001	11.4539	10.7679	10.5599
	PRI	0.6470	0.6736	0.6773	0.6786	0.6799
	VOI	3.2992	3.0480	3.0100	3.0140	3.0342
	GCE	0.4677	0.4244	0.4183	0.4182	0.4202
8	BDE	12.0612	11.8708	11.0190	10.5499	10.1550
	PRI	0.6654	0.6750	0.6861	0.6863	0.6867
	VOI	3.7615	3.6423	3.5292	3.5692	3.5954
	GCE	0.5052	0.4857	0.4646	0.4709	0.4734
10	BDE	11.9241	11.6796	11.1483	10.3658	10.1287
	PRI	0.6719	0.6784	0.6827	0.6869	0.6873
	VOI	3.9876	3.8570	3.8592	3.8481	3.8631
	GCE	0.5118	0.4909	0.4859	0.4811	0.4834
15	BDE	11.5662	11.4719	11.2057	10.6012	10.2308
	PRI	0.6784	0.6809	0.6833	0.6873	0.6894
	VOI	4.4138	4.3356	4.3146	4.2583	4.2618
	GCE	0.5108	0.4975	0.4912	0.4750	0.4781
20	BDE	11.4502	11.3212	11.0895	10.7220	10.3427
	PRI	0.6809	0.6825	0.6842	0.6851	0.6892
	VOI	4.6865	4.6538	4.6317	4.6414	4.5446
	GCE	0.5029	0.4935	0.4868	0.4871	0.4732



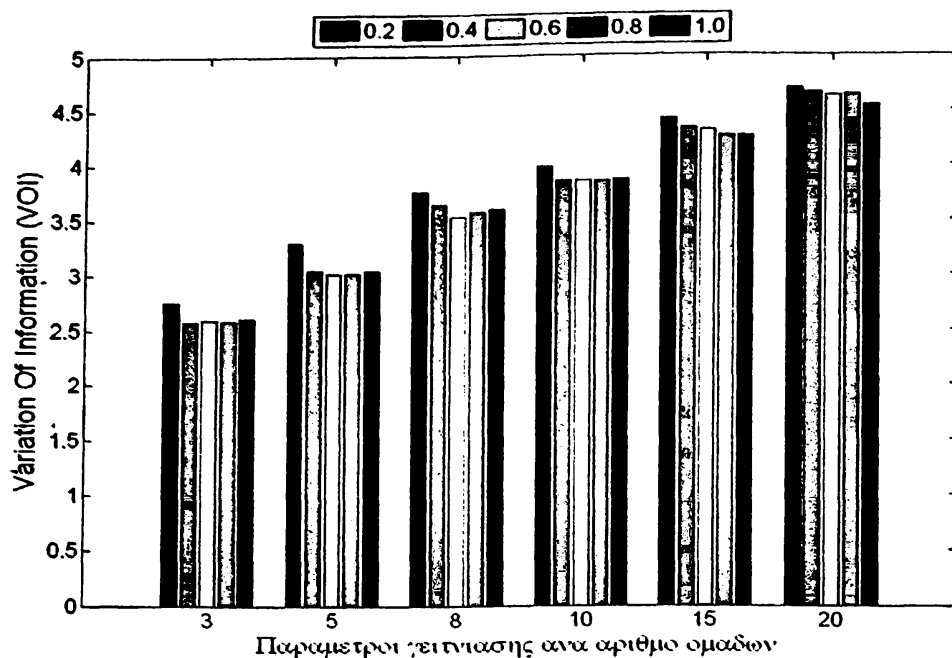


Σχήμα 3.20: Αξιολόγηση των κατατιμήσεων με τη μέθοδο φασματικής ομαδοποίησης με χρήση του πίνακα ομοιότητας και απόκλιση KL. Ο δείκτης BDE. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος.

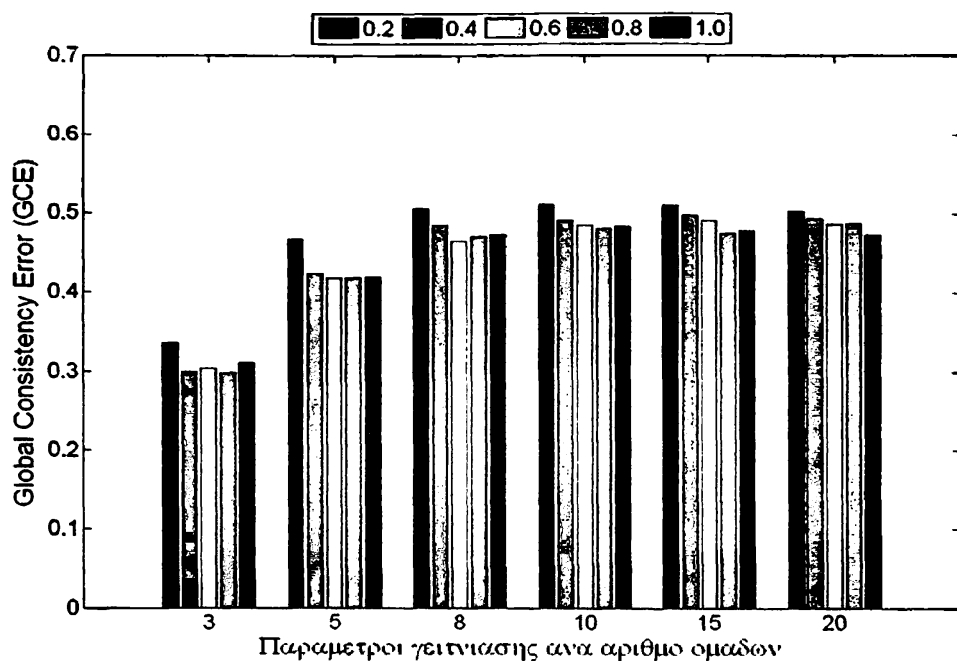


Σχήμα 3.21: Αξιολόγηση των κατατιμήσεων με τη μέθοδο φασματικής ομαδοποίησης με χρήση του πίνακα ομοιότητας και απόκλιση KL. Ο δείκτης PRI. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης a που φαίνονται στο πάνω μέρος του γραφήματος.





Σχήμα 3.22: Αξιολόγηση των κατατιμήσεων με τη μέθοδο φασματικής ομαδοποίησης με χρήση του πίνακα ομοιότητας και απόκλιση KL. Ο δείκτης VOI. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης α που φαίνονται στο πάνω μέρος του γραφήματος.



Σχήμα 3.23: Αξιολόγηση των κατατιμήσεων με τη μέθοδο φασματικής ομαδοποίησης με χρήση του πίνακα ομοιότητας και απόκλιση KL. Ο δείκτης GCE. Στον οριζόντιο άξονα αναφέρονται οι τιμές του αριθμού ομάδων M . Για κάθε τιμή του M , οι στήλες αντιστοιχούν στις τιμές της παραμέτρου γειτνίασης α που φαίνονται στο πάνω μέρος του γραφήματος.



ΚΕΦΑΛΑΙΟ 4

ΚΑΤΑΤΜΗΣΗ ΕΙΚΟΝΑΣ ΜΕ ΜΕΘΟΔΟΥΣ ΕΠΙΒΛΕΠΟΜΕΝΗΣ ΜΑΘΗΣΗΣ

-
- 4.1 Εισαγωγή
 - 4.2 Σημασιολογική κατάτμηση
 - 4.3 Δάση στοιχείων υφής με εννοιολογικά χαρακτηριστικά
 - 4.4 Βάση MSRC21
 - 4.5 Αλγόριθμος
 - 4.6 Πειραματικά Αποτελέσματα
-

4.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται η επιβλεπόμενη μάθηση και η εφαρμογή της για την κατηγοριοποίηση περιοχών σε εικόνες. Στην παράγραφο 4.2 αναλύεται η έννοια της σημασιολογίας στην εικόνα. Στην παράγραφο 4.3 μελετάται η τεχνική των Semantic Texton Forests (STFs) που αποτελεί μία από τις πιο πρόσφατες και μοντέρνες μεθόδους για σημασιολογική κατάτμηση. Στην παράγραφο 4.4 περιγράφεται η Βάση MSRC21 της Microsoft που χρησιμοποιήθηκε στα πειράματα της επιβλεπόμενης μάθησης. Στην παράγραφο 4.5 περιγράφεται αναλυτικά ο βασικός αλγόριθμος επιβλεπόμενης μάθησης. Τέλος, στην παράγραφο 4.6 παρουσιάζονται τα πειραματικά αποτελέσματα της μεθόδου που χρησιμοποιήθηκε.

4.2 Σημασιολογική κατάτμηση

Η κατανόηση της σημασίας των οπτικών δεδομένων είναι εξαιρετικά σημαντική καθώς υπάρχει έξαρση στην ποσότητα των ψηφιακών εικόνων που κυκλοφορούν και καθώς, επιπλέον,



η χρήση υπολογιστών και ειδικά το Διαδίκτυο μετατοπίζεται γρήγορα προς μία κυρίως ει-
κονική αναπαράσταση των δεδομένων. Ωστόσο, το αξιοσημείωτο ποσοστό ποικιλίας εντός
των κατηγοριών (για παράδειγμα διαφορετικά είδη καρεκλών) σε συνδυασμό με την ύπαρξη
ομοιότητας μεταξύ των κατηγοριών (για παράδειγμα παρόμοιες ράτσες σχύλων και γατών)
κάνει αυτό το πρόβλημα αφάνταστα δύσκολο να επιλυθεί. Συγκεκριμένα, η σημασιολογική
κατάτμηση των εικόνων σε συνεχείς περιοχές παρόμοιας ερμηνείας συνδυάζει τις δυσκολίες
της αναγνώρισης αντικειμένων και κατάτμησης εικόνας για να καταλήξει σε ένα πρόβλημα
μεγάλης πολυπλοκότητας αλλά εξαιρετικής αξίας.

Η σημασιολογική κατάτμηση που πραγματοποιούμε προσδίδει σε κάθε εικονοστοιχείο ή
υπερεικονοστοιχείο της εικόνας μία από τις ετικέτες κλάσεων που έχουν καθοριστεί εξαρ-
χής. Επομένως, η αρχική εικόνα χωρίζεται σε περιοχές που αντιστοιχούν σε αντικείμενα
και έννοιες που ο ταξινομητής κατηγοριοποίησε βάσει των ετικετών κλάσεων που έχει μά-
θει. Γενικά υπάρχουν πολλαπλές σημασιολογικές κατατμήσεις μιας εικόνας, ανάλογα με
το λεξικό των αντικειμένων που χρησιμοποιούνται, αλλά κάθε μία δίνει μια αντίληψη της
βαθύτερης σημασίας της εικόνας. Αν και αυτό μπορεί να γίνει από άτομα που μπορούν να
προσθέσουν χειρόγραφες ετικέτες, είναι επίπονο και η τεράστια ποσότητα των εικόνων που
περιλαμβάνονται το κάνει απαγορευτικό. Συνεπώς, η ανάπτυξη πρωτότυπων μεθόδων για
την αυτόματη σημασιολογική κατάτμηση εικόνων είναι απαραίτητη.

Στο σχήμα (4.1) παρουσιάζεται η κατηγοριοποίηση των περιοχών μιας εικόνας μέσω
της σημασιολογικής κατάτμησης της.



Σχήμα 4.1: Αρχική εικόνα και η σημασιολογική της κατάτμηση. Το σχήμα αναπαράχθηκε από
το [6].

4.3 Δάση στοιχείων υψής με εννοιολογικά χαρακτηριστικά

Το δάσος στοιχείων υψής με εννοιολογικά χαρακτηριστικά (STF) [20], [6] είναι ένα μο-
ντέρνο πλαίσιο για αναπαράσταση οπτικών δεδομένων. Αποτελεί ουσιαστικά επέκταση των
τυχαίων δέντρων απόφασης (Randomized Decision Forests ή RDFs) στο πρόβλημα της
σημασιολογικής κατάτμησης και περιγραφής εικόνας. Αρχικά θα αναφέρουμε περιληπτικά
κάποια στοιχεία για τα τυχαία δέντρα απόφασης (RDFs) στο γενικό πλαίσιο της σημα-
σιολογικής κατάτμησης, έπειτα θα περιγράψουμε πως μπορεί να γίνει η εκπαίδευση αυτών
των δέντρων για τη σημασιολογική κατηγοριοποίηση των εικονοστοιχείων της εικόνας και



τέλος θα εξηγήσουμε πώς γίνεται η κατηγοριοποίηση των εικονοστοιχείων μιας τυχαίας εικόνας ελέγχου.

4.3.1 Τυχαία δέντρα απόφασης

Το τυχαίο δέντρο απόφασης (RDF) είναι μία τεχνική μηχανικής μάθησης που μπορεί να χρησιμοποιηθεί για να κατηγοριοποιηθούν ατομικά τα εικονοστοιχεία μιας εικόνας. Βασίζεται στην ιδέα του απλού δέντρου απόφασης, το οποίο χρησιμοποιείται εκτενώς στην εξόρυξη δεδομένων και τη μηχανική μάθηση και το οποίο έχει μία ευρύτερη εφαρμογή σε πολλά πεδία ως εργαλείο σχεδιασμού. Στην πιο απλή του μορφή, ένα δέντρο απόφασης αποτελείται από μία ιεραρχία ερωτήσεων που έχουν ως αποτέλεσμα μία απόφαση. Προκειμένου να πάρουμε μία απόφαση, ξεκινάμε από τη ρίζα και χρησιμοποιούμε την απάντηση στην ερώτηση του κόμβου για να συνεχίσουμε σε ένα από τα παιδιά του, μέχρι να βρεθεί κάποιος κόμβος φύλλο και να παρθεί μια απόφαση. Θέλουμε να χρησιμοποιήσουμε τα δέντρα απόφασης για να κατηγοριοποιήσουμε ατομικά εικονοστοιχεία της εικόνας και, επομένως, οι ερωτήσεις σε κάθε κόμβο θα βασίζονται σε πληροφορία της εικόνας. Συγκεκριμένα, μαθηματικούς συνδυασμούς των τιμών της γειτονιάς του εικονοστοιχείου.

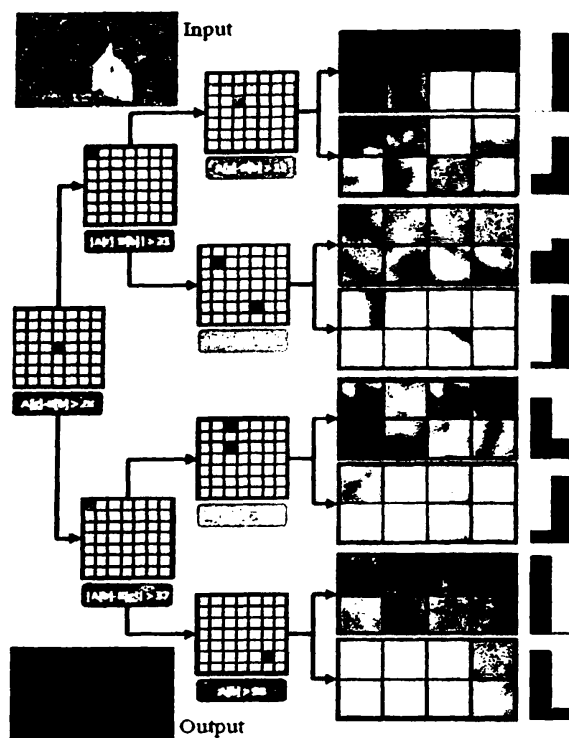
Ένα τυχαίο δέντρο απόφασης είναι ένας συνδυασμός από πολλά διαφορετικά δέντρα απόφασης, κάθε ένα από τα οποία έχει ένα διαφορετικό σύνολο ερωτήσεων για τους κόμβους του και ενδεχομένως μία διαφορετική δομή από τα άλλα δέντρα. Το "τυχαίο" μέρος αφορά τον τρόπο με τον οποίο εκπαιδεύονται τα δέντρα, όπου αντί να επιλέγονται ερωτήσεις χειροκίνητα ή εξαντλητικά από ένα προκαθορισμένο χώρο πιθανών ερωτήσεων, οι ερωτήσεις παράγονται εντελώς τυχαία και διαλέγονται σύμφωνα με μία μετρική επίδοσης, συνήθως το κέρδος πληροφορίας. Στην περίπτωση μας, η βασική μονάδα κατηγοριοποίησης είναι ένα εικονοστοιχείο p και πρέπει να καθορίσουμε την κατηγορία του c . Κάθε δέντρο εκπαιδεύεται σε ένα υποσύνολο των δεδομένων, χρησιμοποιώντας κατηγοριοποιημένα, σε επίπεδο εικονοστοιχείου, δεδομένα εκπαίδευσης. Αυτό οδηγεί σε ένα δέντρο με n κόμβους και l κόμβους φύλλα. Με κάθε κόμβο συνδέεται μια μαθημένη κατανομή κλάσης $P(C_p = c|n)$, όπου c είναι ο χώρος της κατηγορικής τυχαίας μεταβλητής C και C_p είναι η μεταβλητή για ένα εικονοστοιχείο p . Ένα παράδειγμα σημασιολογικού δέντρου απόφασης φαίνεται παρακάτω στο σχήμα (4.2), όπου ένα δέντρο έχει εκπαιδευτεί σε εικόνες με πρόβατα και γρασίδι και μπορεί αποδοτικά να καταταμίσει την εικόνα σύμφωνα με αυτές τις δύο σημασιολογικές κατηγορίες.

Όταν ένα νέο εικονοστοιχείο πρόκειται να κατηγοριοποιηθεί, ολόκληρο το δάσος (RDF) πετυχαίνει μία ακριβή και εύρωστη κατηγοριοποίηση, παίρνοντας το μέσο όρο των κατανομών κλάσεων πάνω σε όλους τους κόμβους φύλλα $L(p) = (l_1, l_2, \dots, l_T)$ που φτάνονται από το εικονοστοιχείο p για όλα τα δέντρα T

$$P(c|L(p)) \propto \sum_{t=1}^T P(c|l_t)P(t) \quad (4.1)$$

Τα RDFs αποτελούν ένα δυναμικό εργαλείο ως ταξινομητές και είναι ένας γρήγορος τρόπος για ομαδοποίηση περιγραφών. Τα δέντρα είναι εύκολο να εκπαιδευτούν και να





Σχήμα 4.2: Απλό σημασιολογικό δέντρο απόφασης. Το σχήμα αναπαράχθηκε από το [6].

μετέπειτα να χρησιμοποιηθούν καθώς μόνο ένα μικρό ποσοστό του δέντρου διασχίζεται για κάθε σημείο δεδομένων, κάνοντας τα ιδανικά για τον υπολογισμό μίας "μαλακής" κατανομής των κατηγοριών για κάθε εικονοστοιχείο. Το αποτέλεσμα του δέντρου απόφασης που εφαρμόζεται σε μια εικόνα είναι είτε ένα ιστόγραμμα φύλλων, δηλαδή ένα ιστόγραμμα με μετρήσεις εικονοστοιχείων πάνω στους κόμβους φύλλα, είτε ένα ιστόγραμμα δέντρου, δηλαδή ένα υπερσύνολο του ιστογράμματος φύλλων που περιλαμβάνει μετρήσεις σε όλα τα κλαδιά. Φυσικά, ένα ιστόγραμμα δέντρου μπορεί να υπολογιστεί από ένα ιστόγραμμα φύλλων αν είναι γνωστή η δομή του δέντρου.

4.3.2 Εκπαίδευση STF

Η εκπαίδευση ενός STF εμπεριέχει τον καθορισμό πολλαπλών παραμέτρων

- αριθμός δέντρων του δάσους ως ισοζύγιο μεταξύ ακρίβειας και ταχύτητας.
- μέγιστο βάθος δέντρου.
- μέγεθος συνόλου δοκιμών ελέγχου σε κάθε κόμβο.
- ποσοστό δεδομένων εκπαίδευσης ανά δέντρο.
- μέγεθος του παραθύρου γειτνίασης για ελέγχους γύρω από κάθε εικονοστοιχείο.
- είδος χρησιμοποιούμενων δοκιμών ελέγχου.



Από αυτές τις παραμέτρους, ο βέλτιστος τύπος και ο αριθμός των δοκιμών ελέγχου εξαρτάται αρκετά ισχυρά από τη φύση του συνόλου δεδομένων και το έργο που πρέπει να πραγματοποιηθεί. Για να ανακαλύψουμε τις καλύτερες παραμέτρους όσον αφορά το έργο της εξαγωγής κατηγορίας για κάθε εικονοστοιχείο της εικόνας, θα πρέπει να σχεδιαστούν μία σειρά πειραμάτων όπου οι παράμετροι αλλάζουν και μετρούνται τα αποτελέσματά τους. Φυσικά, το κόστος σε χρόνο και πόρους θα ήταν απαγορευτικό για την πραγματοποίηση μίας πλήρους εξερεύνησης του χώρου παραμέτρων και, συνεπώς, έχουμε περιορίσει την εξερεύνηση στην επίδραση που μπορεί να έχει η αλλαγή μίας μεταβλητής, ενώ οι άλλες παραμένουν σταθερές.

Τα δεδομένα εκπαίδευσης αποτελούνται από ένα σύνολο εικονοστοιχείων P που δειγματοληπτούνται από τις εικόνες εκπαίδευσης με ρυθμό ανά 4 εικονοστοιχεία σε κάθε κατεύθυνση. Αυτό γίνεται προκειμένου να μειωθεί ο χρόνος που απαιτείται για την κατασκευή του δέντρου. Ωστόσο, μετά τη δημιουργία του δέντρου χρησιμοποιούνται όλα τα εικονοστοιχεία των δεδομένων εκπαίδευσης για να συμπληρωθεί η πληροφορία των κόμβων ως εξής

$$P(C|n) = \frac{H_n[c]}{\sum_c H_n[c]} \quad \forall \text{ κόμβο } n \quad (4.2)$$

όπου $H_n[c]$ είναι ο αριθμός των εικονοστοιχείων της κατηγορίας c που πέρασαν διαμέσω του κόμβου n κατά τη διάρκεια της εκπαίδευσης. Με αυτόν τον τρόπο αποφεύγεται οποιοδήποτε "υπερταίριασμα" προκαλείται από την κατασκευή του δέντρου σε ένα υποσύνολο δεδομένων.

Μετά τη δειγματοληψία των εικονοστοιχείων, κατασκευάζουμε το δέντρο χωρίζοντας αναδρομικά το P σε δύο υποσύνολα $P_{left} \cup P_{right} = P$ βασιζόμενοι σε μία επιλεγμένη δοκιμή ελέγχου. Το P_{left} στέλνεται στο αριστερό παιδί και το P_{right} στέλνεται στο δεξί παιδί και η διαδικασία επαναλαμβάνεται μέχρι να συναντηθεί μία συνθήκη τερματισμού. Η δοκιμή ελέγχου, βάσει της οποίας θα πραγματοποιηθεί ο διαχωρισμός του P , επιλέγεται από ένα σύνολο δοκιμών ελέγχου και είναι αυτή που έχει ως αποτέλεσμα το μεγαλύτερο αναμενόμενο κέρδος σε πληροφορία σε σχέση με τις κατηγορίες των κόμβων. Το κέρδος υπολογίζεται ως εξής

$$\Delta E = -\frac{|P_{left}|}{|P|} E(P_{left}) - \frac{|P_{right}|}{|P|} E(P_{right}) \quad (4.3)$$

όπου $E(P)$ είναι η εντροπία Shannon των κλάσεων στο σύνολο P .

Αφού κατασκευαστεί το δέντρο, υπολογίζεται η κατανομή (4.2) που αποτελεί μία καλή εκτίμηση της κατανομής πιθανότητας $P(c|n)$ για κάθε κόμβο. Στο [6] υπάρχουν περισσότερες λεπτομέρειες τόσο για την επιλογή των κατάλληλων παραμέτρων εκπαίδευσης όσο και για την κατασκευή του STF με εναλλακτικές τεχνικές.

4.3.3 Κατηγοριοποίηση STF

Η αξιοποίηση του STF για την απόδοση κατηγοριών στα εικονοστοιχεία μίας εικόνας ελέγχου πραγματοποιείται με τη χρήση της (4.1) σε όλα τα φύλλα των δέντρων του δάσους που καταλήγει το εικονοστοιχείο. Στο [6] εκπαιδεύτηκαν 10 δέντρα σε ξεχωριστά



υποσύνολα του συνόλου δεδομένων της βάσης MSRC21, η οποία περιγράφεται στην ενότητα (4.4). Έπειτα δημιουργήθηκαν 10 δάση από 5 δέντρα, τα οποία αποτελούν μεταθέσεις των 10 δέντρων που είχαν εκπαιδευτεί. Η απόφαση κατηγοριοποίησης λαμβάνεται βάσει και των 10 δασών που χρησιμοποιήθηκαν.

Θεωρώντας ένα πίνακα "σύγχυσης" M ως εξής

$$M[i, j] = |\{p : p \in P_R, G(p) = c_i, \operatorname{argmax}_c P(c|L_p) = c_j\}| \quad (4.4)$$

όπου P_R το σύνολο των εικονοστοιχείων της εικόνας R , οι αναφερόμενες τιμές ακρίβειας είναι η μέση ακρίβεια

$$m = \frac{1}{|\{c_i\}|} \sum_{c_i} a_{c_i} \quad (4.5)$$

όπου c_i η i -οστή κλάση και $a_{c_i} = \frac{M[i, i]}{\sum_j M[i, j]}$ καθώς και η συνολική ακρίβεια

$$a = \frac{\sum_i M[i, i]}{\sum_i \sum_j M[i, j]} \quad (4.6)$$

σε 10 δοκιμές σημασιολογικής κατάτμησης που πραγματοποιήθηκαν στα δάση, αφού είχαν εκπαιδευτεί και είχαν καθοριστεί οι παράμετροι τους. Στο [6] υπάρχουν αναλυτικά τα πειραματικά αποτελέσματα και οι σημασιολογικές κατηγοριοποιήσεις που παράχθηκαν.

4.4 Βάση MSRC21

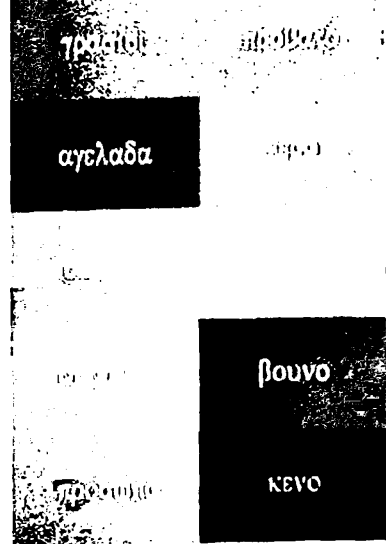
Η βάση MSRC21 της Microsoft περιέχει ένα μεγάλο πλήθος εικόνων που μπορούν να χρησιμοποιηθούν στην περιοχή της σημασιολογικής κατάτμησης εικόνας. Οι εικόνες που περιλαμβάνονται έχουν επιπλέον καταταμηθεί με χειρόγραφες ετικέτες. Αποτελείται από 591 εικόνες που υπάρχουν σε έγχρωμη μορφή και οι οποίες έχουν χαρακτηριστεί με χειρόγραφες ετικέτες που αφορούν 21 κατηγορίες. Επιπλέον υπάρχει μία "κενή" κατηγορία, που αφορά εικονοστοιχεία τα οποία δεν μπορούν να χαρακτηρισθούν και να αποδοθούν σε κάποια από τις 21 κατηγορίες.

Στο σχήμα (4.3) φαίνονται κάποια παραδείγματα εικόνων της Βάσης MSRC21 και οι ετικέτες των κατηγοριών τους. Στην εργασία μελετήθηκαν 10 κατηγορίες (γρασίδι, πρόβατο, καφέ αγελάδα, μαύρη αγελάδα, νερό, δέντρο, λουλούδι, ουρανός, βουνό και πρόσωπο), από τις οποίες οι δύο κατηγορίες που αφορούν την αγελάδα (καφέ αγελάδα και μαύρη αγελάδα) έχουν συνενωθεί οπτικά σε μία. Συνεπώς, στο σχήμα φαίνονται οι ετικέτες των 10 κλάσεων (μαζί με την "κενή") που χρησιμοποιήθηκαν στα πειραματικά αποτελέσματα.

4.5 Αλγόριθμος

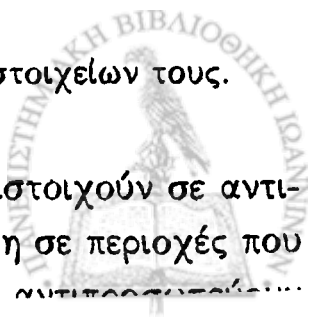
Στόχος του αλγορίθμου επιβλεπόμενης μάθησης σε επίπεδο εικόνας είναι η κατηγοριοποίηση των υπερεικονοστοιχείων της σε κάποια από τις 10 κλάσεις. Ιδεατά επιθυμούμε οι περιοχές





Σχήμα 4.3: Εικόνες της βάσης MSRC21 και οι κατηγορίες των εικονοστοιχείων τους.

ου θα προκύψουν μετά την κατάταξη των υπερεικονοστοιχείων να αντιστοιχούν σε αντι-
είμενα και έννοιες της εικόνας. Δηλαδή, η εικόνα να είναι κατετημημένη σε περιοχές που
πορούν να χαρακτηριστούν ανεξάρτητες μεταξύ τους και οι οποίες να αντιπροσωπεύουν



ξεχωριστά οντότητες του πραγματικού κόσμου. Συμπερασματικά, επιθυμούμε να προσεγγίσουν όσο το δυνατόν καλύτερα κάθε πιθανή χειρόγραφη κατάτμηση της εικόνας από κάποιο τυχαίο άτομο, το οποίο θα αντιστοιχούσε κάθε οπτικά διακριτή περιοχή της εικόνας με μία διαφορετική ετικέτα.

Αρχικά, πριν την εφαρμογή του αλγορίθμου, θα πρέπει να μοντελοποιήσουμε τις 10 κλάσεις ώστε να μπορέσουμε να τις μάθουμε. Συγκεκριμένα, χρησιμοποιούνται δύο στρατηγικές

1. Μοντελοποιούμε κάθε κλάση με κανονική κατανομή $N(\mu, \Sigma)$ με μέση τιμή μ το μέσο όρο των τιμών των MRF χαρακτηριστικών των εικονοστοιχείων της στις εικόνες εκπαίδευσης και πίνακα συμμεταβλητότητας Σ τον πίνακα συνδιακύμανσης που αντιστοιχεί σε αυτές τις τιμές. Συνεπώς, η εκπαίδευση συνίσταται στην εύρεση της κανονικής κατανομής $N(\mu, \Sigma)$ για κάθε κλάση.
2. Μοντελοποιούμε κάθε κλάση με όλες τις τιμές των MRF χαρακτηριστικών των εικονοστοιχείων της στις εικόνες εκπαίδευσης. Έπειτα, τροφοδοτούμε τις τιμές αυτές με την αντίστοιχη ετικέτα κλάσης σε ένα ταξινομητή SVM [2], ο οποίος αναλαμβάνει την εκμάθηση και παράγει το μοντέλο εκπαίδευσης.

Παρακάτω ακολουθεί συνοπτικά η περιγραφή του αλγορίθμου (6) στο πλαίσιο της κατηγοριοποίησης των υπερεικονοστοιχείων μιας εικόνας σε κάποια από τις 10 κλάσεις, ενώ αρχικά είχε χρησιμοποιηθεί η πρώτη στρατηγική εκμάθησης.

Εναλλακτικά, μπορούμε να τροποποιήσουμε το βήμα 4 του αλγορίθμου (6), όπου αντί για την απόκλιση KL, χρησιμοποιούμε την απόσταση κοντινότερου γείτονα 1-NN

$$d_{ij}(\mu_i, \mu_j) = (\mu_i - \mu_j)^T (\mu_i - \mu_j)$$

Επίσης, εφαρμόζουμε και μηχανές διανυσμάτων υποστήριξης (SVMs). Ακολουθεί συνοπτικά η περιγραφή του αλγορίθμου (7) στο πλαίσιο της κατηγοριοποίησης των υπερεικονοστοιχείων μιας εικόνας σε κάποια από τις 10 κλάσεις, ενώ αρχικά είχε χρησιμοποιηθεί η δεύτερη στρατηγική εκμάθησης.

Επίσης, μπορούμε να πραγματοποιήσουμε σημασιολογική κατάτμηση σε επίπεδο εικονοστοιχείου, δηλαδή να αναθέτουμε κάθε φορά μία κλάση σε κάθε εικονοστοιχείο. Σε αυτή την περίπτωση, τροποποιούμε τα βήματα των αλγορίθμων (6) και (7), θέτοντας 'εικονοστοιχείο' στη θέση του 'υπερεικονοστοιχείο'. Η μόνη διαφορά, πλέον, είναι ότι κάθε εικονοστοιχείο μοντελοποιείται απλά με την τιμή των MRF χαρακτηριστικών του. Αυτό γιατί κάθε εικονοστοιχείο μπορεί να θεωρηθεί εκφυλισμένη περίπτωση υπερεικονοστοιχείου (αρκεί να θεωρήσουμε $\Sigma = I$, όπου I ο μοναδιαίος πίνακας, και η απόκλιση KL μετατρέπεται σε απόσταση κοντινότερου γείτονα).

Μία άλλη παραλλαγή των αλγορίθμων (6) και (7) είναι να ακολουθήσουμε διαφορετική τακτική όσον αφορά την απόφαση για την κατηγοριοποίηση των υπερεικονοστοιχείων. Συγκεκριμένα, κατηγοριοποιούμε αρχικά κάθε εικονοστοιχείο ενός υπερεικονοστοιχείου και, έπειτα, αποδίδουμε στο υπερεικονοστοιχείο την κλάση που εμφανίζεται σε μεγαλύτερο ποσοστό στα εικονοστοιχεία του.



Αλγόριθμος 6 Αλγόριθμος επιβλεπόμενης μάθησης με ταξινομητή KL

- 1 Δοθείσης εικόνας L , χρησιμοποιούμε τα δοσμένα υπερεικονοστοιχεία της. Αυτά είχαν προκύψει εφαρμόζοντας τον αλγόριθμο Neut.
- 2 Εξάγουμε τα MRF χαρακτηριστικά της εικόνας, τα οποία κρατάμε χωρίς μείωση διάστασης και χωρίς κανονικοποίηση.
- 3 Μοντελοποιούμε κάθε υπερεικονοστοιχείο με κανονική κατανομή $N(\mu, \Sigma)$ με μέση τιμή μ το μέσο όρο των τιμών των MRF χαρακτηριστικών των εικονοστοιχείων του και πίνακα συμμεταβλητότητας Σ τον πίνακα συνδιακύμανσης που αντιστοιχεί σε αυτές τις τιμές.
- 4 Κατηγοριοποιούμε κάθε υπερεικονοστοιχείο i αποδίδοντας του την κλάση j που έχει τη μικρότερη απόκλιση KL

$$d_{ij}(N(\mu_i, \Sigma_i), N(\mu_j, \Sigma_j)) = (\mu_i - \mu_j)^T (\mu_i - \mu_j) + \text{Tr} \left(\Sigma_i + \Sigma_j - 2(\Sigma_i \Sigma_j)^{\frac{1}{2}} \right)$$

από την κατανομή που το χαρακτηρίζει.

- 5 Κατασκευάζουμε τη σημασιολογική εικόνα θέτοντας ως ετικέτα όλων των εικονοστοιχείων κάθε υπερεικονοστοιχείου αυτή της κλάσης που του ανατέθηκε, αντίστοιχα, στο βήμα 4.

Αλγόριθμος 7 Αλγόριθμος επιβλεπόμενης μάθησης με ταξινομητή SVM

- 1 Δοθείσης εικόνας L , χρησιμοποιούμε τα δοσμένα υπερεικονοστοιχεία της. Αυτά είχαν προκύψει εφαρμόζοντας τον αλγόριθμο Neut.
- 2 Εξάγουμε τα MRF χαρακτηριστικά της εικόνας, τα οποία κρατάμε χωρίς μείωση διάστασης και χωρίς κανονικοποίηση.
- 3 Μοντελοποιούμε κάθε υπερεικονοστοιχείο με τη μέση τιμή του μ που είναι ο μέσος όρος των τιμών των MRF χαρακτηριστικών των εικονοστοιχείων του.
- 4 Κατηγοριοποιούμε κάθε υπερεικονοστοιχείο i αποδίδοντας του την κλάση j που επιστρέφει ο ταξινομητής SVM με τη μεγαλύτερη πιθανότητα.

$$\text{κλάση}(i) = \text{argmax}_j P(\text{κλάση} = j|i) \quad (4.7)$$

- 5 Κατασκευάζουμε τη σημασιολογική εικόνα θέτοντας ως ετικέτα όλων των εικονοστοιχείων κάθε υπερεικονοστοιχείου αυτή της κλάσης που του ανατέθηκε, αντίστοιχα, στο βήμα 4.



Οι διαφορετικές παραλλαγές των αλγορίθμων (6) και (7) παρουσιάζονται με εικόνες και πίνακες αξιολόγησης κατάτμησης στα πειραματικά αποτελέσματα (4.6).

4.6 Πειραματικά Αποτελέσματα

Παρακάτω παρατίθενται διάφορες εικόνες της βάσης MSRC21 και οι σημασιολογικές κατατμήσεις τους καθώς και πίνακες με τους δείκτες αξιολόγησης της ποιότητας των σημασιολογικών κατατμήσεων.

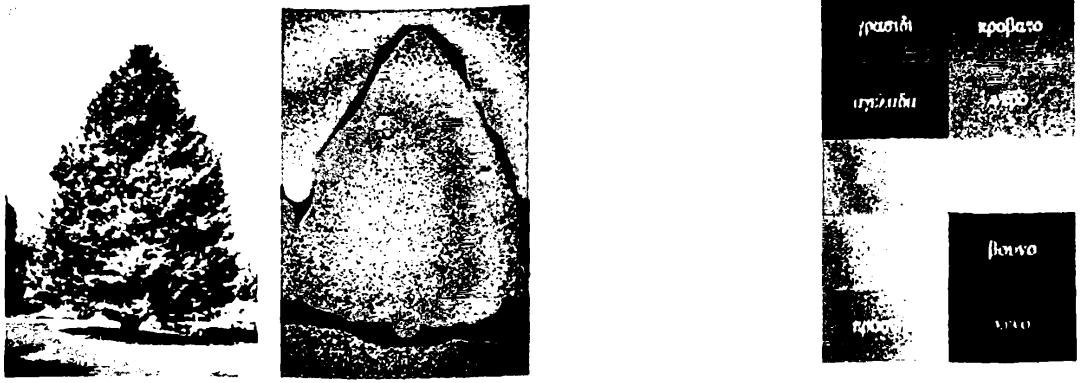
Συγκεκριμένα, παρουσιάζονται εικόνες που προκύπτουν από την εφαρμογή των αλγορίθμων (6) και (7) τόσο σε επίπεδο υπερεικονοστοιχείου όσο και σε επίπεδο εικονοστοιχείου. Χρησιμοποιείται η απόσταση κοντινότερου γείτονα σε επίπεδο υπερεικονοστοιχείου, η οποία είναι γενικά καλύτερη από την απόκλιση KL. Επίσης, εφαρμόζεται και η στρατηγική της κατηγοριοποίησης ενός υπερεικονοστοιχείου με την κλάση της πλειοψηφίας των εικονοστοιχείων του. Επιπλέον, παρουσιάζονται δείκτες αξιολόγησης της ποιότητας των σημασιολογικών κατατμήσεων. Υπολογίζεται η ακρίβεια της σημασιολογικής κατάτμησης για κάθε κλάση ξεχωριστά καθώς και η μέση ακρίβεια (4.5) και η ολική ακρίβεια (4.6) για το σύνολο των 9 κλάσεων (η αγελάδα αποτελεί μία ενιαία κλάση). Σημειώνεται ότι

- Η εκπαίδευση των 10 κλάσεων που χρησιμοποιήθηκαν πραγματοποιήθηκε με κατάλληλες εικόνες εκπαίδευσης, προκειμένου να υπάρχει εμφανής διαχωρισμός των κλάσεων. Για παράδειγμα, η κλάση πρόβατο θεωρήθηκε ότι περιέχει εικονοστοιχεία με άσπρο χρώμα, η κλάση λουλούδι θεωρήθηκε ότι περιέχει εικονοστοιχεία με κόκκινο χρώμα κ.λ.π.
- Χρησιμοποιήθηκαν αποκλειστικά τα MRF χακτηριστικά (2.2) και όχι τα SIFT χακτηριστικά (2.3). Πειραματικά, τα MRF χακτηριστικά αποδίδουν καλύτερα την πληροφορία υψής, καθώς εμπεριέχουν χρωματική πληροφορία που αφορά την περιοχή γύρω από το εικονοστοιχείο. Αντίθετα, τα SIFT χαρακτηριστικά περιέχουν πληροφορία τοπικής παραγωγού της περιοχής γύρω από το εικονοστοιχείο και είναι δύσκολο να υπάρξει σαφής διαχωρισμός των κλάσεων. Αυτό συμβαίνει καθώς οι περισσότερες από τις 10 κλάσεις αναφέρονται σε αντικείμενα και έννοιες που οπτικά είναι ομοιογενείς περιοχές με μικρές χρωματικές αλλαγές. Συνεπώς, η τοπική παράγωγος παίρνει πολύ μικρές τιμές για αυτές τις κατηγορίες και, επομένως, δε μας βοηθά να τις διαχωρίσουμε.
- Για την εκπαίδευση του ταξινομητή SVM (2.7.2) χρησιμοποιήθηκε αρχικά η τεχνική 'cross-validation' για την εκμάθηση των παραμέτρων C της 'C-Support Vector' ταξινόμησης [2] και γ της ακτινικής συνάρτησης βάσης πυρήνα

$$\exp(-\gamma * \|u - v\|^2)$$

Έπειτα, χρησιμοποιούμε όλα τα δεδομένα εκπαίδευσης για να δημιουργήσουμε το μοντέλο με τις καλύτερες παραμέτρους C και γ που βρήκαμε.





(a)

(b)

(c)



(d)

(e)

(f)

(g)

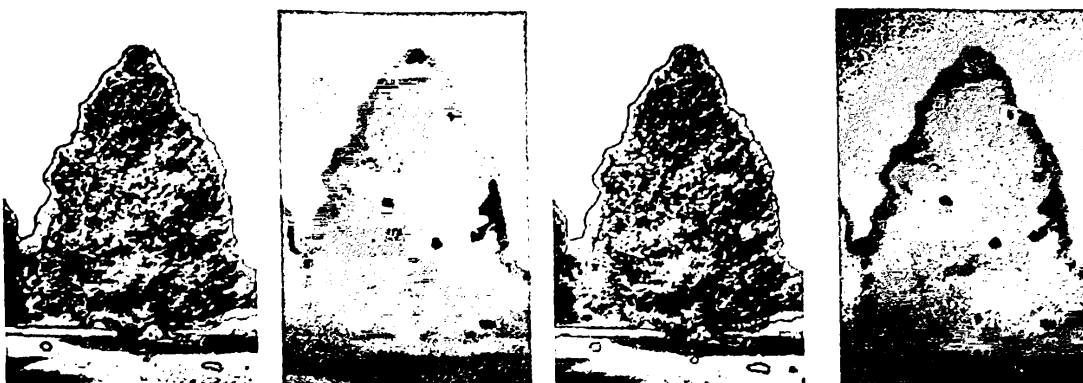


(h)

(i)

(j)

(k)



(l)

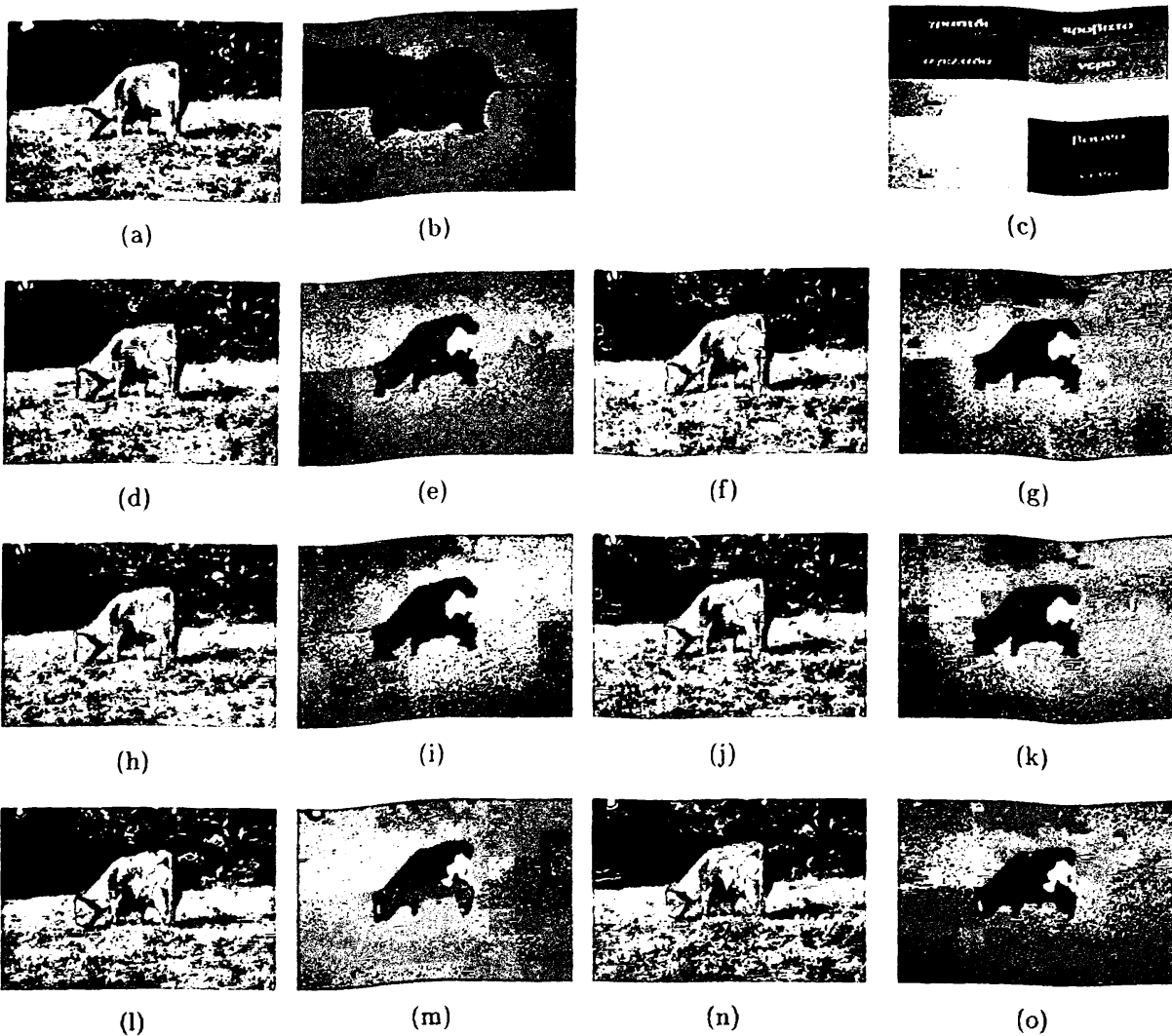
(m)

(n)

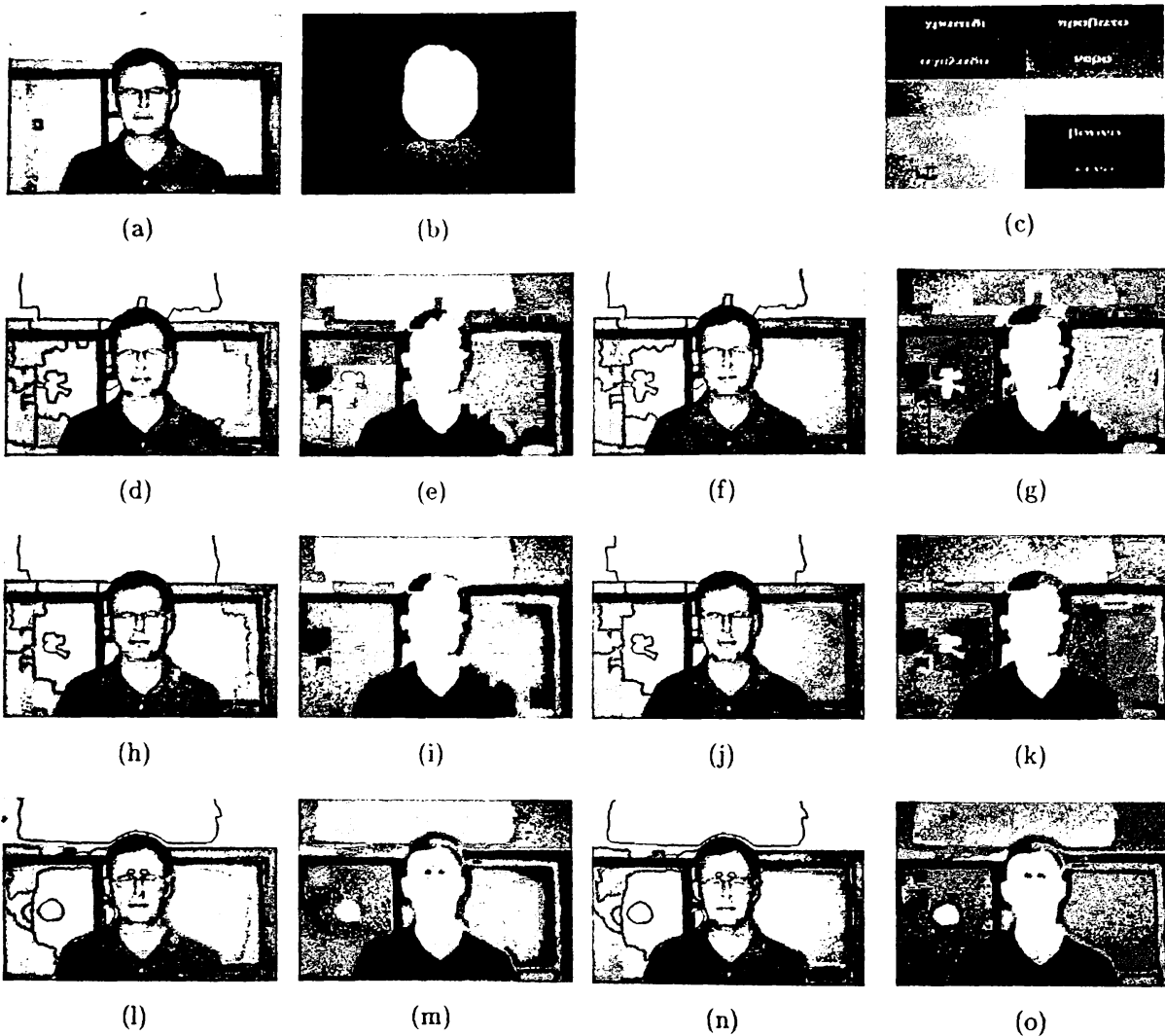
(o)

α 4.4: Σημασιολογική κατάτμηση. (a) Εικόνα. (b) Κατηγορίες. (c) Αντιστοιχίες με χρώματα. (d)-(e) 1-NN σε υπερεικονοστοιχεία. (f)-(g) SVM σε υπερεικονοστοιχεία. (h)-(i) 1-NN με ψηφοφορία σε υπερεικονοστοιχεία. (j)-(k) SVM με ψηφοφορία σε υπερεικονοστοιχεία. (l)-(m) 1-NN σε εικονοστοιχεία. (n)-(o) SVM σε εικονοστοιχεία.

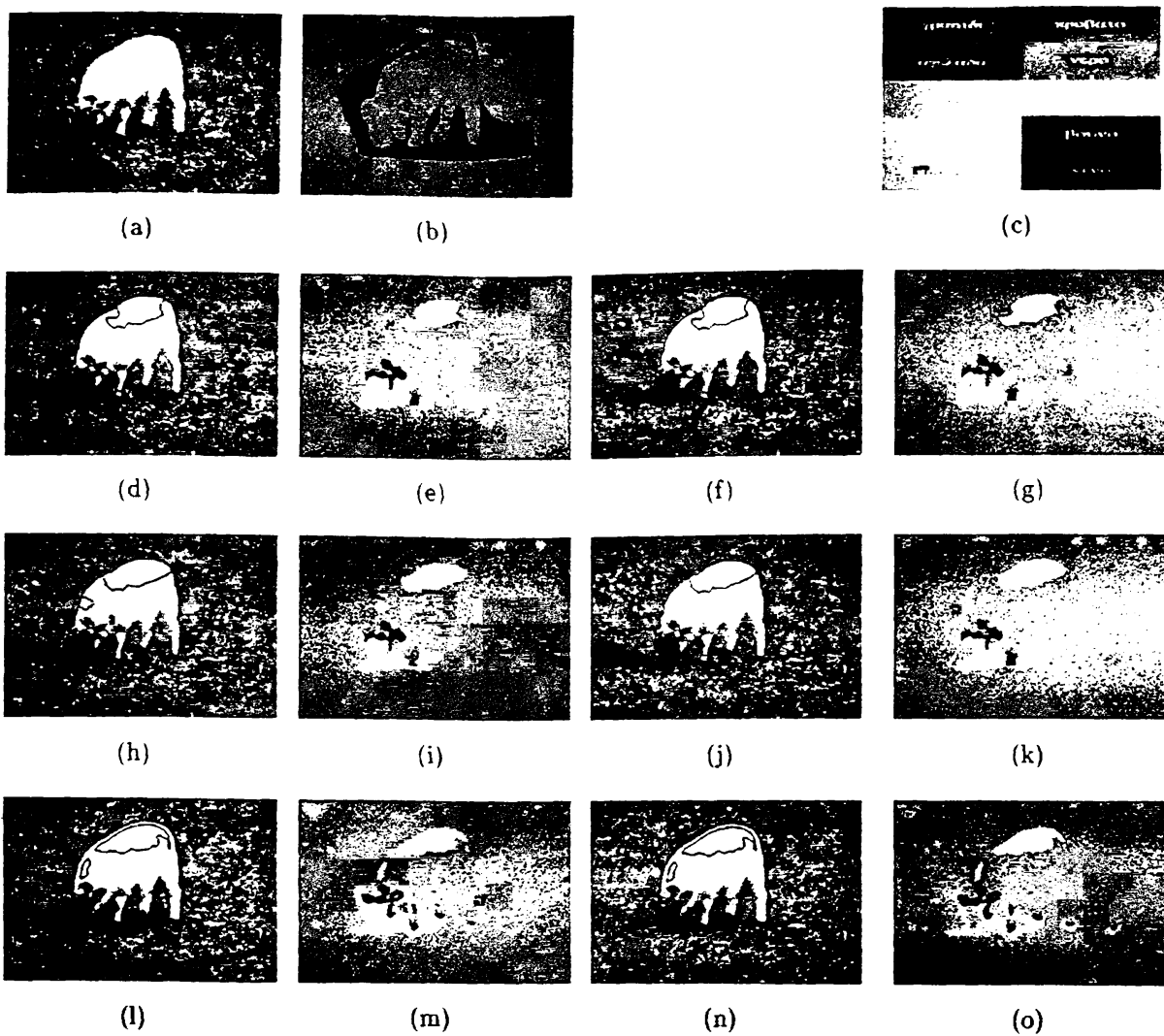




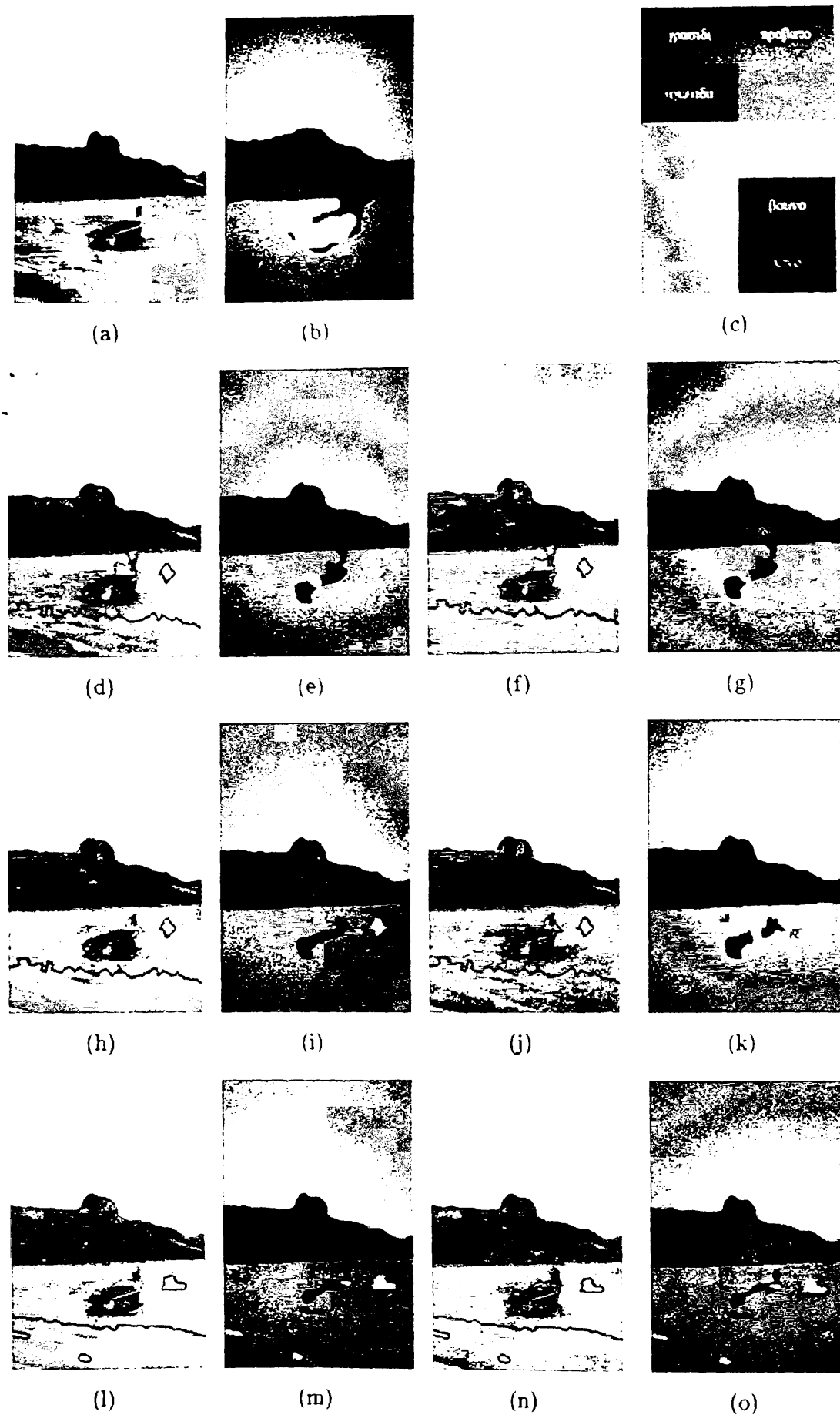
Σχήμα 4.5: Σημασιολογική κατάτμηση. (a) Εικόνα. (b) Κατηγορίες. (c) Αντιστοίχιση κατηγοριών με χρώματα. (d)-(e) 1-NN σε υπερεικονοστοιχεία. (f)-(g) SVM σε υπερεικονοστοιχεία. (h)-(i) 1-NN με ψηφοφορία σε υπερεικονοστοιχεία. (j)-(k) SVM με ψηφοφορία σε υπερεικονοστοιχεία. (l)-(m) 1-NN σε εικονοστοιχεία. (n)-(o) SVM σε εικονοστοιχεία.



Σχήμα 4.6: Σημασιολογική κατάτμηση. (a) Εικόνα. (b) Κατηγορίες. (c) Αντιστοίχιση κατηγοριών με χρώματα. (d)-(e) 1-NN σε υπερεικονοστοιχεία. (f)-(g) SVM σε υπερεικονοστοιχεία. (h)-(i) 1-NN με ψηφοφορία σε υπερεικονοστοιχεία. (j)-(k) SVM με ψηφοφορία σε υπερεικονοστοιχεία. (l)-(m) 1-NN σε εικονοστοιχεία. (n)-(o) SVM σε εικονοστοιχεία.



Σχήμα 4.7: Σημασιολογική κατάτμηση. (a) Εικόνα. (b) Κατηγορίες. (c) Αντιστοίχιση κατηγοριών με χρώματα. (d)-(e) 1-NN σε υπερεικονοστοιχεία. (f)-(g) SVM σε υπερεικονοστοιχεία. (h)-(i) 1-NN με ψηφοφορία σε υπερεικονοστοιχεία. (j)-(k) SVM με ψηφοφορία σε υπερεικονοστοιχεία. (l)-(m) 1-NN σε εικονοστοιχεία. (n)-(o) SVM σε εικονοστοιχεία.



1.8: Σημασιολογική κατάτμηση. (a) Εικόνα. (b) Κατηγορίες. (c) Αντιστοιχίες με χρώματα. (d)-(e) 1-NN σε υπερεικονοστοιχεία. (f)-(g) SVM σε υπερεικονοστοιχεία. (h)-(i) 1-NN με ψηφοφορία σε υπερεικονοστοιχεία. (j)-(k) SVM με ψηφοφορία σε υπερεικονοστοιχεία. (l)-(m) 1-NN σε εικονοστοιχεία. (n)-(o) SVM σε εικονοστοιχεία.



Πίνακας 4.1: Αξιολόγηση των σημασιολογικών κατατμήσεων με ταξινομητή 1-NN και απόκλιση KL. Συμβολίζουμε με 'sup's' τα υπερεικονοστοιχεία και με 'pixels' τα εικονοστοιχεία.

ταξινομητής κατηγορίες	1-NN sup's	1-NN ψηφοφορία sup's	1-NN pixels	KL sup's
ολική ακρίβεια	0.6690	0.6728	0.6710	0.6884
μέση ακρίβεια	0.4934	0.4950	0.4919	0.5080
γρασίδι	0.8167	0.8222	0.8207	0.8135
πρόβατο	0.4415	0.4412	0.4411	0.3616
αγελάδα	0.5718	0.5791	0.5751	0.7005
νερό	0.7642	0.7673	0.7670	0.7970
δέντρο	0.3659	0.3678	0.3655	0.4883
λουλούδι	0.6605	0.6736	0.6576	0.7201
ουρανός	0.6604	0.6641	0.6671	0.6498
βουνό	0.2562	0.2519	0.2479	0.1315
πρόσωπο	0.3969	0.3829	0.3773	0.4174

Πίνακας 4.2: Αξιολόγηση των σημασιολογικών κατατμήσεων με ταξινομητή SVM. Συμβολίζουμε με 'sup's' τα υπερεικονοστοιχεία και με 'pixels' τα εικονοστοιχεία.

ταξινομητής κατηγορίες	SVM sup's	SVM ψηφοφορία sup's	SVM pixels
ολική ακρίβεια	0.6646	0.6671	0.6657
μέση ακρίβεια	0.4862	0.4853	0.4832
γρασίδι	0.8145	0.8192	0.8179
πρόβατο	0.4468	0.4446	0.4450
αγελάδα	0.5700	0.5737	0.5695
νερό	0.7666	0.7690	0.7696
δέντρο	0.3646	0.3682	0.3648
λουλούδι	0.5818	0.5803	0.5706
ουρανός	0.6591	0.6624	0.6655
βουνό	0.2589	0.2521	0.2501
πρόσωπο	0.3999	0.3831	0.3787



ΚΕΦΑΛΑΙΟ 5

ΕΠΙΛΟΓΟΣ

5.1 Επίλογος

5.1 Επίλογος

Στην εργασία ασχοληθήκαμε με την απλή και τη σημασιολογική κατάτμηση εικόνας. Στο κεφάλαιο 2 παρουσιάστηκε το θεωρητικό υπόβαθρο πάνω στο οποίο στηρίζονται οι μέθοδοι επιβλεπόμενης και μη επιβλεπόμενης μάθησης των κεφαλαίων 3 και 4.

Στο κεφάλαιο 3 παρουσιάστηκαν οι αλγόριθμοι μη εποπτευόμενης μάθησης και η εφαρμογή τους για την ανίχνευση ομοιογενών περιοχών σε εικόνες. Ένα σημαντικό θέμα για συζήτηση και μελλονική ενασχόληση είναι η περαιτέρω βελτίωση της κατάτμησης. Για παράδειγμα, να επιδιωχθεί η λεπτομερής μελέτη οριακών υπερεικονοστοιχείων γειτονικών ομοιογενών περιοχών προκειμένου να αποφασιστεί αν θα πρέπει να αλλάξουν ή να διατηρήσουν την περιοχή τους. Επίσης, να μελετηθούν περιοχές που σχηματικά βρίσκονται στο εσωτερικό άλλων περιοχών και πιθανώς αποτελούν τμήμα της ίδιας ευρύτερης οντότητας.

Στο κεφάλαιο 4 παρουσιάστηκαν οι αλγόριθμοι εποπτευόμενης μάθησης και η εφαρμογή τους για εντοπισμό συγκεκριμένων κατηγοριών σε εικόνες. Ένα ζήτημα για μελλοντική ερευνητική δραστηριότητα είναι η σε βάθος μελέτη της έννοιας της σημασιολογικής κατάτμησης, η οποία εμπεριέχει και την αλληλοσυσχέτιση των κατηγοριών στο ευρύτερο εννοιολογικό πλαίσιο. Για παράδειγμα, η πιθανότητα να υπάρχει πρόβατο δίπλα από γρασίδι είναι μεγαλύτερη από την πιθανότητα να υπάρχει βιβλίο δίπλα από γρασίδι. Επομένως, μπορεί να επεκταθεί η σημασιολογική κατάτμηση προκειμένου να συμπεριλαμβάνει κάποια πρότερη γνώση για τη "συνύπαρξη" των κατηγοριών και, συνεπώς, να μπορεί να κατηγοριοποιεί καλύτερα υπερεικονοστοιχεία που δεν είναι εύκολο να διακριθούν αποκλειστικά μέσω των χαρακτηριστικών τους λόγω της μεγάλης ποικιλομορφίας των κατηγοριών (για παράδειγμα άσπρα βιβλία και άσπρα πρόβατα).



ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [2] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001.
- [3] C. Fowlkes, S. Belongie, F. Chung, and J. Malik. Spectral grouping using the nystrom method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):214–225, 2004.
- [4] J. Freixenet, X. Munoz, D. Raba, J. Marti, and X. Cufi. Yet another survey on image segmentation: Region and boundary information integration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 408–422, 2002.
- [5] R. C. Gonzalez and R. E. Woods. *Digital Image Processing, Third Edition*. Prentice-Hall, Inc., 2006.
- [6] M. Johnson. *Semantic Segmentation and Image Search*. PhD thesis, University of Cambridge, April 2008.
- [7] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2169–2178, 2006.
- [8] D. G. Lowe. Distinctive Image Features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [9] U. V. Luxburg, M. Belkin, O. Bousquet, and Pertinence. A tutorial on spectral clustering. *Statistics and Computing*, 17(4), 2007.
- [10] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 416–423, 2001.
- [11] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 416–423, July 2001.
- [12] D. R. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):530–549, 2004.



- [13] M. Meila. Comparing clusterings: An axiomatic view. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 577–584. ACM Press, 2005.
- [14] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.
- [15] G. Mori. Guiding model search using segmentation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1417–1423, 2005.
- [16] C. Pantofaru and M. Hebert. A comparison of image segmentation algorithms. Technical report, The Robotics Institute, Carnegie Mellon University, 2005.
- [17] X. Ren and J. Malik. Learning a classification model for segmentation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 10–17, 2003.
- [18] J. Shi, C. Fowlkes, D. Martin, and E. Sharon. Graph based image segmentation tutorial. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.
- [19] J. Shi and J. Malik. Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
- [20] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2008.
- [21] P. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison Wesley, 2006.
- [22] S. Theodoridis and K. Koutroumbas. *Pattern Recognition, Fourth Edition*. Academic Press, Inc., 2009.
- [23] M. Varma and A. Zisserman. Texture classification: Are filter banks necessary? In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 691–698, 2003.
- [24] A. Yang, J. Wright, S. Sastry, and Y. Ma. Unsupervised segmentation of natural images via lossy data compression. *Computer Vision and Image Understanding*, 110(2):212–225, 2008.



ΒΙΟΓΡΑΦΙΚΟ

Ο Χρήστος Παππάς γεννήθηκε στα Ιωάννινα το 1987. Αποφοίτησε το 2005 από το 4^ο ενιαίο λύκειο Ιωαννίνων. Οι βασικές σπουδές πραγματοποιήθηκαν στο Τμήμα Πληροφορικής του πανεπιστημίου Ιωαννίνων από όπου και αποφοίτησε το 2009. Συνέχισε για Μεταπτυχιακές σπουδές στο ίδιο Ίδρυμα και ασχολήθηκε με το πεδίο της κατάκτησης εικόνας. Για το διάστημα των προπτυχιακών σπουδών διατέλεσε και υπότροφος του Ιδρύματος Κρατικών Υποτροφιών.



