



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΕΙΚΟΝΩΝ, ΒΙΝΤΕΟ ΚΑΙ ΠΟΛΥΜΕΣΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Τεχνολογίες Γνώσης και Πρόσβαση σε
Οπτικοακουστικές Πληροφορίες

Διδακτορική Διατριβή του
Αθανάσιου Αθανασιάδη
Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Ε.Μ.Π. (2003)

Αθήνα, Νοέμβριος 2009



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΕΙΚΟΝΩΝ, ΒΙΝΤΕΟ ΚΑΙ ΠΟΛΥΜΕΣΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Τεχνολογίες Γνώσης και Πρόσβαση σε Οπτικοακουστικές Πληροφορίες

Διδακτορική Διατριβή του

Αθανάσιου Αθανασιάδη

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Ε.Μ.Π. (2003)

Συμβουλευτική Επιτροπή: Στέφανος Κόλλιας
Ανδρέας-Γεώργιος Σταφυλοπάτης
Παναγιώτης Τσανάκας

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 20η Νοεμβρίου 2009.

...
Σ. Κόλλιας	Α.-Γ. Σταφυλοπάτης	Π. Τσανάκας
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

...
Π. Μαραγκός	Ν. Μήτρου	Γ. Στάμου
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Λέκτορας Ε.Μ.Π.

...

Μ. Ζερβάκης
Καθηγητής Πολυτεχνείου Κρήτης

Αθήνα, Νοέμβριος 2009



Η παρούσα διδακτορική διατριβή πραγματοποιήθηκε στα πλαίσια του προγράμματος ΠΕΝΕΔ-2003, της Γενικής Γραμματείας Έρευνας και Τεχνολογίας. Το πρόγραμμα συγχρηματοδοτήθηκε κατά 80% από την Ευρωπαϊκή Ένωση και κατά 20% από το Ελληνικό Δημόσιο.

This Ph.D. thesis was supported by grant PENED-2003 of the Greek Ministry of Development-GSRT and was co-financed by E.U.-European Social Fund (80%) and National Resources (20%).

...

Αθανάσιος Αθανασιάδης

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

Copyright © Αθανάσιος Αθανασιάδης, 2009.

Με επιφύλαξη παντός δικαιώματος - All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Πρόλογος

Η μελέτη και η εργασία για την εκπόνηση ενός διδακτορικού είναι αέναη και κοπιώδης, αλλά ταυτόχρονα ιδιαίτερα ενδιαφέρουσα και εξόχως δημιουργική. Η επίτευξη του στόχου αυτού, μου έδωσε προσωπικά πολύ μεγάλη ικανοποίηση, τόσο για την ολοκλήρωση και το τελικό επιστημονικό αποτέλεσμα, όσο και για τις εμπειρίες που αποκόμισα και τους ανθρώπους που γνώρισα όλα αυτά τα χρόνια. Αυτή η περίοδος αποτέλεσε το πιο γόνιμο και ζωντανό κομμάτι της ζωής μου, έως σήμερα, και αποτελεί οδηγό για τη στάση ζωής που θέλω να κρατήσω από εδώ και στο εξής.

Είχα την τύχη να εργαστώ υπό τις καλύτερες δυνατές συνθήκες που θα μπορούσα να έχω, στο περιβάλλον του εργαστηρίου Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων, και σε αυτό έχει συμβάλει καθοριστικά ο διευθυντής του εργαστηρίου, καθηγητής Στέφανος Κόλλιας. Όλα αυτά τα χρόνια ο κ. Κόλλιας φροντίζει προσωπικά, με τη θετική ενέργεια που τον διακρίνει πάντα, ώστε το εργαστήριο να αποτελεί τον κατάλληλο χώρο για μελέτη και παραγωγή δημιουργικής σχέψης, αλλά παράλληλα και έναν ευχάριστο μικρόκοσμο του οποίου έχω την τιμή να είμαι μέλος.

Θερμές ευχαριστίες οφείλω και στα άλλα δύο μέλη της τριμελούς επιτροπής, καθηγητές Ανδρέα Σταφυλοπάτη και Παναγιώτη Τσανάκα, οι οποίοι ήταν πρόθυμοι να με βοηθήσουν όσες φορές χρειάστηκε. Ευχαριστώ επίσης και τα υπόλοιπα μέλη της εξεταστικής επιτροπής, καθηγητές Πέτρο Μαραγκό, Νικόλαο Μήτρου, Γιώργο Στάμου και Μιχάλη Ζερβάκη, για το ενδιαφέρον τους στο αντικείμενο της διατριβής μου και τα θετικά τους σχόλια. Ιδιαίτερη μνεία θα ήθελα να κάνω στον Δρ. Γιάννη Αβρίθη, ο οποίος αποτέλεσε για μένα πρότυπο εργατικότητας, δημιουργικής και συγχροτημένης σχέψης και η επιρροή του ήταν μεγάλη σε ερευνητικό, αλλά και σε ένα ευρύτερο επίπεδο. Θα ήθελα επίσης να ευχαριστήσω τους ερευνητές και τους υποψήφιους διδάκτορες του εργαστηρίου. Με ορισμένους διασταυρωθήκαμε ερευνητικά, μοιραστήκαμε χαρές και απογοητεύσεις, αγωνίες και ξενύχτια για την επίτευξη των εκάστοτε στόχων μας. Με πολλούς άλλους μπορεί να μην συνεργάστηκα ερευνητικά, αλλά συνέβαλαν καθοριστικά στην επιτυχή περάτωση του διδακτορικού μου με τη συνεχή τους υποστήριξη, τις συζητήσεις μας επί παντός επιστητού και τους ευρύτερους κοινούς προβληματισμούς μας.

Τέλος, θέλω να ευχαριστήσω τους φίλους μου, οι οποίοι πάντα με στήριξαν στα δύσκολα και με τους οποίους επίσης μοιράστηκα όλες τις χαρές μου. Χάρη σε αυτούς δεν ένιωσα ποτέ ότι αποκόπηκα από τις υπόλοιπες διαστάσεις της ζωής, ιδιαίτερα τους τελευταίους, δύσκολους μήνες κατά την ολοκλήρωση του διδακτορικού μου. Ξέρουν πολύ καλά πόσο σημαντική είναι για μένα η έννοια της παρέας και η φιλία τους σημαίνει πολλά για μένα. Τέλος, είναι προφανές, και για τους δύο μας, ότι ένα 'ευχαριστώ' δεν αρκεί για να ανταποδώσει την υποστήριξη και τα εφόδια που φρόντισε να μου δώσει η μητέρα μου. Ελπίζω να καταφέρω να φανώ αντάξιος των προσδοκιών όλων όσοι με πιστεύουν και με αγαπάνε· είναι ένα βαρύ αλλά πολύ ευχάριστο φορτίο.

Θάνος Αθανασιάδης
Αθήνα, Νοέμβριος 2009

Περιεχόμενα

Πρόλογος	v
Περιεχόμενα	vii
Κατάλογος Σχημάτων	xi
Κατάλογος Πινάκων	xiii
Περίληψη	xv
Abstract	xvii
1 Εισαγωγή	1
1.1 Αντικείμενο και Κίνητρο Διατριβής	1
1.2 Διάρθρωση Κειμένου	3
2 Αναπαράσταση Γνώσης	7
2.1 Εισαγωγή	7
2.2 Ασαφή Σύνολα	8
2.2.1 Ορισμοί και συμβολισμοί	8
2.2.2 Ασαφείς Τελεστές	10
2.2.3 Ασαφείς Σχέσεις	11
2.3 Θεωρία Γράφων	11
2.3.1 Αναπαράσταση Εικόνας με Γράφο	14
2.4 Οντολογικό Μοντέλο Γνώσης	16
2.4.1 Η Γλώσσα Αναπαράστασης Γνώσης RDF	17
2.4.2 Η Γλώσσα Αναπαράστασης Γνώσης RDFs	19
2.4.3 Η Γλώσσα Αναπαράστασης Γνώσης OWL	20
2.4.4 Ασαφής Επέκταση της Γλώσσας OWL	21
3 Εξαγωγή Σημασιολογικής Περιγραφής με Χρήση Τεχνολογιών Γνώσης	23
3.1 Εισαγωγή	23
3.2 Επισκόπηση Μεθοδολογίας	24
3.3 Οντολογική Υποδομή Γνώσης	25
3.3.1 Οντολογία Οπτικών Περιγραφέων	26
3.3.2 Οντολογία Δομής Πολυμέσων	26
3.3.3 Οντολογία Πυρήνα και Χωρο-χρονικές Επεκτάσεις	27
3.3.4 Οντολογία Θεματικού Πεδίου	28

3.3.5	Προσέγγιση με Πρότυπα Στιγμιότυπα	28
3.4	Εξαγωγή και Αποθήκευση Χαρακτηριστικών	29
3.4.1	Αρχική Κατάτμηση.....	29
3.4.2	Εξαγωγή και Ταίριασμα Χαρακτηριστικών	29
3.4.2.1	Περιγραφέας Κύριων Χρωμάτων	30
3.4.2.2	Περιγραφέας Ομοιογενούς Υφής	30
3.4.2.3	Περιγραφέας Σχήματος με βάση την Περιοχή	31
3.4.3	Ανάκτηση Πρότυπων Στιγμιότυπων	31
3.4.4	Αρχικοποίηση Σχεσιακού Γράφου Χαρακτηριστικών <i>ARG</i>	33
3.5	Κατασκευή Σημασιολογικών Μεταδεδομένων.....	35
4	Σημασιολογική Κατάτμηση Εικόνων και Βίντεο	39
4.1	Κατάτμηση Εικόνων	39
4.1.1	Ταξινόμια Αλγορίθμων Κατάτμησης	40
4.1.2	Αλγόριθμος Κατάτμησης RSST.....	42
4.1.3	Αλγόριθμος Κατάτμησης Watershed	43
4.1.4	Αξιολόγηση Αλγορίθμων Κατάτμησης	44
4.2	Σημασιολογική Επέκταση Περιοχών	46
4.2.1	Επισκόπηση Προτεινόμενης Προσέγγισης	47
4.2.2	Αλγόριθμος Σημασιολογικού Πλημμυρισμού	48
4.2.3	Αλγόριθμος Σημασιολογικού RSST	51
4.3	Σημασιολογική Κατάτμηση Βίντεο	53
4.3.1	Επισκόπηση Μεθοδολογίας	54
4.3.2	Αναπαράσταση Βίντεο με Γράφο	55
4.3.3	Ανάλυση ενός BoF	56
4.3.3.1	Επιλογή Καρέ	56
4.3.3.2	Υπολογισμός Σημασιολογικής και Συντακτικής Περιγραφής Όγκου	57
4.3.3.3	3Δ Σημασιολογική Επέκταση	57
4.3.4	Ανάλυση Διαδοχικών BoFs.....	58
4.3.4.1	Ταίριασμα Όγκων Διαδοχικών BoFs.....	58
4.3.4.2	Ενημέρωση και Διάδοση Εννοιών	60
4.4	Οπτικό Εννοιολογικό Πλαίσιο	62
4.5	Πειραματικά Αποτελέσματα σε Εικόνες	63
4.5.1	Προετοιμασία Πειραμάτων και Διαδικασία Αξιολόγησης	63
4.5.2	Ενδεικτικά Αποτελέσματα και Συμπεράσματα	65
4.5.3	Εννοιολογικό Πλαίσιο	68
4.5.4	Συνολικά Αποτελέσματα	71
4.6	Πειραματικά Αποτελέσματα σε Ακολουθίες Βίντεο	73
4.7	Συμπεράσματα	76
5	Σημασιολογική Ερμηνεία και Συλλογιστική: Οπτική Λογική	77
5.1	Εισαγωγή	77
5.2	Προσδιορισμός του Προβλήματος	78
5.3	Αξιωματικός Σκελετός και Ερμηνεία της Οπτικής Λογικής.....	81
5.3.1	Μοντέλο Γνώσης Οπτικής Πληροφορίας.....	81
5.3.2	Περιγραφικές Λογικές για Αναπαράσταση Οπτικής Πληροφορίας	83

5.3.3	Κατανόηση Περιεχομένου και Συλλογιστική.....	84
5.4	Από τους Οπτικούς Περιγραφείς στις Έννοιες-Διεγέρτες.....	87
5.4.1	Διεγέρτες Χρώματος.....	87
5.4.2	Διεγέρτες Υφής.....	89
5.4.3	Διεγέρτες Σχήματος.....	90
5.5	Αρχικά Αποτελέσματα.....	91
5.6	Συναφείς Εργασίες.....	93
5.7	Συμπεράσματα.....	95
6	Επίλογος	97
6.1	Συμπεράσματα.....	97
6.2	Συνεισφορά και Δημοσιεύσεις.....	98
6.3	Μελλοντικές Επεκτάσεις.....	100
A	Κατάλογος Απόδοσης Όρων	103
A.1	Κατάλογος Αγγλικών Όρων.....	103
A.2	Κατάλογος Ελληνικών Όρων.....	107
A.3	Αχρωνύμια.....	111
B	Κατάλογος δημοσιεύσεων	113
B.1	Περιοδικά.....	113
B.2	Κεφάλαια Βιβλίων.....	114
B.3	Συνέδρια.....	114
B.4	Τεχνικές Αναφορές.....	116
	Βιβλιογραφία	117
	Βιογραφικό Σημείωμα	129

Κατάλογος Σχημάτων

1.1	Σχηματικό διάγραμμα ανάλυσης περιεχομένου με βοήθεια γνώσης.	3
2.1	Παραδείγματα επίπεδων γράφων.	12
2.2	Σχέση υπογράφου και επαγόμενου υπογράφου.	13
2.3	Εξαγωγή ελάχιστου συνεκτικού δένδρου από σταθμισμένο γράφο.	14
2.4	Παράδειγμα αναπαράστασης κατάτμησης S μίας εικόνας με ένα σχεσιακό γράφο χαρακτηριστικών.	15
3.1	Υποδομή και διασύνδεση οντολογιών.	26
3.2	Ιεραρχία οντολογίας οπτικών περιγραφέντων.	27
3.3	Αρχιτεκτονική ανάλυσης πολυμέσων με βοήθεια γνώσης.	33
3.4	Αποτελέσματα σε εικόνες από το θεματικό πεδίο “παραλία”.	36
3.5	Αποτελέσματα σε εικόνες από το θεματικό πεδίο “φόρμουλα 1”.	37
3.6	Αποτελέσματα σε εικόνες από το θεματικό πεδίο “τένις”.	37
4.1	Κατάτμηση εικόνας από διαφορετικούς χρήστες (1ο παράδειγμα από τη βάση εικόνων του Berkeley).	40
4.2	Κατάτμηση εικόνας από διαφορετικούς χρήστες (2ο παράδειγμα από τη βάση εικόνων του Berkeley).	41
4.3	Διάγραμμα ροής του αλγορίθμου κατάτμησης RSST.	42
4.4	Μονοδιάστατο παράδειγμα κατάτμησης watershed.	44
4.5	Κατάταξη μεθοδολογιών αξιολόγησης αλγορίθμων κατάτμησης εικόνων.	45
4.6	Σχηματική αναπαράσταση του σημασιολογικού πλημμυρισμού εικόνας.	50
4.7	Διάγραμμα λειτουργίας σημασιολογικής κατάτμησης βίντεο.	54
4.8	Χωρική και χρονική αποδόμηση ενός BoF B_i	55
4.9	Ταίριασμα όγκων διαδοχικών BoFs.	60
4.10	Συγχώνευση δύο BoFs.	61
4.11	Εικόνες από το σύνολο δεδομένων αλήθειας.	65
4.12	Κατάτμηση εικόνας με χρήση των αλγορίθμων RSST, SW και $SRSST$ (1ο παράδειγμα).	66
4.13	Κατάτμηση εικόνας με χρήση των αλγορίθμων RSST, SW και $SRSST$ (2ο παράδειγμα).	67
4.14	Κατάτμηση εικόνας με χρήση των αλγορίθμων RSST, SW και $SRSST$ (3ο παράδειγμα).	68
4.15	Σημασιολογική κατάτμηση εικόνων και αναπροσαρμογή των βαθμών εμπιστοσύνης των εντοπισμένων εννοιών μέσω ανατροφοδότησης της αναγνώρισης.	69
4.16	Σημασιολογική κατάτμηση και εννοιολογικό πλαίσιο.	70
4.17	Μελέτη επίδρασης θορύβου εισόδου στην τελική ταξινόμηση.	72

4.18	Σημασιολογική κατάτμηση βίντεο (1ο παράδειγμα).....	74
4.19	Σημασιολογική κατάτμηση βίντεο (2ο παράδειγμα).....	75
5.1	Κατάτμηση εικόνας με βάση τη χρωματική πληροφορία.	79
5.2	Αρχιτεκτονική του προτεινόμενου συστήματος οπτικής λογικής.....	81
5.3	Συναρτήσεις συμμετοχής για τις έννοιες-διεγέρτες χρώματος.	89
5.4	Κατάτμηση εικόνας τη χρονική στιγμή t	93

Κατάλογος Πινάκων

2.1	Ασαφείς τελεστές.....	11
4.1	Βαθμοί συμμετοχής κάθε έννοιας για τέσσερις γειτονικές περιοχές. ...	67
4.2	Υπολογισμός βαρών των ακμών του υπογράφου G_{SEA}	68
4.3	Βαθμοί συμμετοχής κάθε έννοιας πριν και μετά τον αλγόριθμο εννοιολογικού πλαισίου για εικόνα από το θεματικό πεδίο “παραλία”... ..	70
4.4	Σύγκριση επιδόσεων ανά έννοια των προτεινόμενων αλγορίθμων για το θεματικό πεδίο “παραλία”.....	71
4.5	Αξιολόγηση σημασιολογικής κατάτμησης βίντεο.	76
5.1	Σημασιολογία εννοιών και ρόλων.	85
5.2	Ορισμός βαθμών συμμετοχής διεγερτών υφής.....	90
5.3	Δείγμα της βάσης γνώσης Σ	92
5.4	Ασαφείς ισχυρισμοί από ένα δείγμα του $ABox$	94

Περίληψη

Το αντικείμενο της διατριβής αυτής είναι η ενσωμάτωση τεχνολογιών γνώσης στην ανάλυση και την περιγραφή εικόνων και βίντεο. Η σύμπραξη των τεχνολογιών γνώσης με τεχνικές όρασης υπολογιστών στοχεύει στην βελτίωση της κατανόησης οπτικού περιεχομένου, μέσω της αξιοποίησης πρότερης γνώσης σε αλγόριθμους σημασιολογικής ανάλυσης εικόνων και βίντεο. Πιο συγκεκριμένα, μελετάμε το πρόβλημα της κατάτμησης εικόνων και βίντεο και προτείνουμε καινοτόμες τεχνικές εντοπισμού, εξαγωγής, αναγνώρισης και παρακολούθησης αντικειμένων βάση σημασιολογικών και οπτικών κριτηρίων. Προτείνουμε μία προσέγγιση *σημασιολογικής κατάτμησης*, εμπλουτίζοντας τεχνικές επέκτασης περιοχών με σημασιολογικά χαρακτηριστικά προκειμένου να αντιμετωπίσουμε προβλήματα που οφείλονται στην αδυναμία περιγραφής σημασιολογικών οντοτήτων αποκλειστικά με οπτικά χαρακτηριστικά. Επιπλέον, προτείνουμε ένα δομημένο πλαίσιο γνώσης, το οποίο αποκαλούμε *οπτική λογική*, βασισμένο στις περιγραφικές λογικές και τις ασαφείς επεκτάσεις τους, για να συνδέσουμε οπτικά χαρακτηριστικά με έννοιες που αποτελούν το λεξιλόγιο ενός θεματικού πεδίου. Χρησιμοποιώντας ένα σύνολο αξιωμάτων και αξιοποιώντας υπηρεσίες συλλογιστικής, εξάγουμε συμπεράσματα για τις πιθανές εννοιολογικές ερμηνείες που επιδέχονται τμήματα ή το σύνολο της εικόνας.

□

Abstract

The main research area of this Ph.D thesis, in a broad sense, is the integration of knowledge technologies into the analysis and description of multimedia. Knowledge technologies can aid computer vision tasks towards the improvement of the understanding of visual content, by exploiting a priori knowledge in algorithms of semantic image and video analysis. More specifically, we examine the problem of image and video segmentation and we propose novel techniques for the detection, extraction, recognition and tracking of objects, based on semantic and visual criteria. We propose a *semantic segmentation* approach, which enhances region growing algorithms with semantic characteristics, in order to deal with problems that raise from the shortcoming of describing semantic entities by visual characteristics exclusively. Moreover, we propose a structured knowledge framework, which we call *visual logics*, based on description logics and their fuzzy extensions, to link visual data with concepts that form the vocabulary of a domain. We use a set of axioms and a reasoning engine to infer possible semantic interpretation of parts or the whole of an image.

□

Κεφάλαιο 1

Εισαγωγή

1.1 Αντικείμενο και Κίνητρο Διατριβής

Είναι αδιαμφισβήτητο γεγονός ότι ζούμε στην εποχή της πληροφορίας, η οποία χαρακτηρίζεται από τη δυνατότητα για εύκολη πρόσβαση στην πληροφόρηση και τη γνώση. Μέχρι πρόσφατα, η ψηφιακή πληροφορία περιοριζόταν κατά το μεγαλύτερο ποσοστό της σε κείμενο. Η πρόοδος στις τεχνολογίες υλισμικού και τηλεπικοινωνιών έχει οδηγήσει σε μια συνεχή και ταχύτατη αύξηση του όγκου του παραγόμενου πολυμεσικού (έντυπου, ηχητικού και οπτικού) περιεχομένου, το οποίο εκτός από συλλογές και αρχεία οργανισμών και ινστιτούτων, είναι πλέον πραγματικότητα και για τους απλούς χρήστες. Η παραγωγή και η διαχείριση πολυμεσικού περιεχομένου έχει απλοποιηθεί τόσο πολύ, που απλοί χρήστες δημιουργούν προσωπικές συλλογές εικόνων και βίντεο και τις μοιράζονται μεταξύ τους, παράγοντας έναν τεράστιο όγκο δεδομένων. Ενδεικτικά, σε πρόσφατη έρευνα αναφέρθηκε ότι στο δικτυακό χώρο youtube συλλέγονται 20 ώρες βίντεο κάθε λεπτό, ενώ υπηρεσίες διαχείρισης και διανομής εικόνων όπως το flickr και το photobucket έχουν συγκεντρώσει από τους χρήστες τους περισσότερες από 10^{10} εικόνες.

Όλο αυτό το παραγόμενο υλικό αξιοποιείται σε ένα ευρύ φάσμα εφαρμογών που εκτείνεται από υπηρεσίες ψυχαγωγίας, τουρισμού, πολιτισμού, εξ αποστάσεως εκμάθησης μέχρι το ηλεκτρονικό εμπόριο και την τηλεϊατρική. Το σημαντικότερο πρόβλημα που προκύπτει, είναι η αποτελεσματική διαχείριση του τεράστιου αυτού όγκου δεδομένων, καθώς η χρησιμότητα όλων αυτών των εφαρμογών καθορίζεται πλέον, σχεδόν αποκλειστικά, από τις δυνατότητες που παρέχονται για εύκολη αναζήτηση, ταχεία πρόσβαση και φιλική προς το χρήστη παρουσίαση του σχετικού περιεχομένου. Το πολυμεσικό περιεχόμενο αποτελεί μια μεγάλη πρόκληση από την άποψη της αποθήκευσης, της μετάδοσης, της δεικτοδότησης και της ανάκτησης. Για να αντιμετωπισθούν τέτοιες προκλήσεις δεν αρκεί να αναπτυχθεί απλώς γρηγορότερο υλισμικό ή να σχεδιαστούν περισσότεροι αλγόριθμοι εξαγωγής και επεξεργασίας χαρακτηριστικών από το σήμα καθαυτό, αλλά απαιτείται μια βαθύτερη κατανόηση του πολυμεσικού υλικού στο σημασιολογικό επίπεδο. Κάτι τέτοιο απαιτεί την αναγνώριση αντικειμένων και γεγονότων που εμφανίζονται στο πολυμεσικό περιεχόμενο με τρόπο αντίστοιχο με την αντίληψη που έχει και την ερμηνεία που τους αποδίδει ο άνθρωπος.

Η απαίτηση για αυτόματη σημασιολογική κατανόηση πολυμεσικού περιεχομένου δεν είναι καινούρια· έχει αποτελέσει μία από τις σημαντικότερες ερευνητικές προκλήσεις των επιστημονικών πεδίων της υπολογιστικής όρασης και της τεχνητής νοημοσύνης και είναι γνωστή με τον όρο *σημασιολογικό κενό* (semantic gap). Ένας πολύ

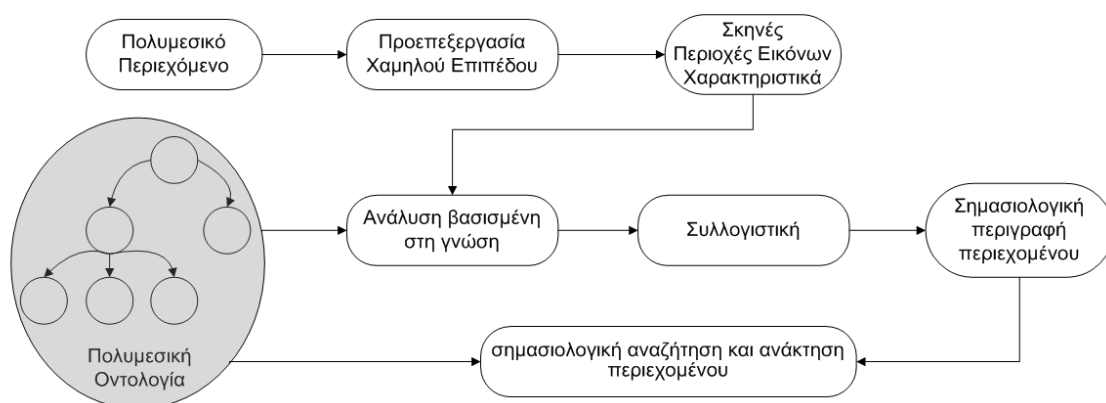
διαδεδομένος ορισμός του σημασιολογικού κενού στο δόθηκε από τους Smeulders et al. [118]: *“The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.”* Σε ελεύθερη απόδοση, το σημασιολογικό κενό αποτελεί την έλλειψη σύμπτωσης μεταξύ της πληροφορίας που μπορεί να εξαχθεί από τα οπτικά δεδομένα και της ερμηνείας που επιδέχονται τα ίδια δεδομένα για ένα χρήστη σε ορισμένη κατάσταση. Ο όρος “κενό” προκύπτει από την αναντιστοιχία των χαρακτηριστικών που εξάγονται αυτόματα από αλγόριθμους επεξεργασίας σήματος (τα οποία έχουν αριθμητική μορφή), με την ερμηνεία που τους δίνει και μπορεί να αντιληφθεί ο άνθρωπος. Καθώς σκοπός της σημασιολογικής ανάλυσης είναι η εξαγωγή περιγραφής του περιεχομένου ώστε να γίνεται κατανοητή από τον άνθρωπο, είναι αναγκαίο να συνυπολογιστεί το ανθρώπινο γνωσιακό μοντέλο.

Η παραπάνω διαπίστωση οδηγεί στο συμπέρασμα ότι είναι απαραίτητη η ενσωμάτωση γνώσης στη διαδικασία της σημασιολογικής ανάλυσης. Η γνώση γύρω από έννοιες, οντότητες, αντικείμενα και συμβάντα που εμφανίζονται στο περιεχόμενο, καθώς και γύρω από τις σχέσεις μεταξύ των εννοιών αυτών μπορεί να αποτυπωθεί με τρόπο ρητό ή υπονοούμενο, να περιγράφει πρότερη εμπειρία ή να είναι προσαρμοζόμενη, αλλά σε κάθε περίπτωση είναι απαραίτητη η αναπαράστασή της σε μια τυποποιημένη μορφή και κυρίως αναγνωρίσιμη από υπολογιστές. Πρόσφατα έχουν πραγματοποιηθεί εξελίξεις στην κατεύθυνση της αναπαράστασης και εξαγωγής γνώσης. Ειδικότερα, στα πλαίσια της ανάπτυξης του σημασιολογικού ιστού [14], δόθηκε έμφαση στην καθιέρωση τυπικών γλωσσών αναπαράστασης γνώσης, προκειμένου να δοθεί νόημα στην πληροφορία που είναι διαθέσιμη στον παγκόσμιο ιστό και να επιτραπεί η καλύτερη αναζήτηση, ανάκτηση και αξιοποίησή της.

Οι οντολογίες αποτελούν ένα σύγχρονο και ευρέως διαδεδομένο φορμαλισμό αποθήκευσης και αναπαράστασης γνώσης και εξασφαλίζουν τη διαλειτουργικότητα μεταξύ ετερογενών ευφυών συστημάτων και ανθρώπων. Για την κατασκευή, την κωδικοποίηση και τη χρήση μίας οντολογίας, είναι απαραίτητη μία τυπική γλώσσα αναπαράστασης γνώσης. Το πιο σημαντικό χαρακτηριστικό μίας γλώσσας αναπαράστασης οντολογιών είναι η εκφραστικότητά τους· όσο πιο εκφραστική είναι μία γλώσσα, τόσο πιο εύκολο και απλό είναι να αναπαραστήσει κάτι. Είναι όμως γνωστό το πρόβλημα του συμβιβασμού μεταξύ της σημασιολογικής εκφραστικότητας και της πολυπλοκότητας της συλλογιστικής για την εξαγωγή συμπερασμάτων. Η δυνατότητα παραγωγής νέας γνώσης, με χρήση μηχανισμών συλλογιστικής, αποτελεί έναν από τους κύριους στόχους των επιμέρους τεχνικών αναπαράστασης γνώσης και μπορεί να αξιοποιηθεί σε συστήματα σημασιολογικής ανάλυσης πολυμεσικού περιεχομένου.

Στο σχήμα 1.1 απεικονίζεται το γενικό σχηματικό διάγραμμα σημασιολογικής ανάλυσης περιεχομένου με χρήση τεχνολογιών γνώσης, καθώς και οι επιμέρους διαδικασίες και η αλληλεπίδραση αυτών. Οι απαιτήσεις ενός τέτοιου συστήματος είναι να δέχεται ως είσοδο οποιαδήποτε μορφή πολυμεσικού περιεχομένου (κείμενο, ήχος, ομιλία, εικόνα, βίντεο, συνθετικά γραφικά, κ.ά.) και να δίνει στο χρήστη τη δυνατότητα για σημασιολογική αναζήτηση, ανάκτηση και πλοήγηση μέσω ερωτημάτων σε εννοιολογικό επίπεδο.

Αρχικά, το υλικό υφίσταται προεπεξεργασία χαμηλού επιπέδου, όπως είναι για παράδειγμα η αφαίρεση ψηφιακού θορύβου και ανάλογα με το είδος του πολυμεσικού υλικού μπορεί να περιλαμβάνει αύξηση της χρωματικής αντίθεσης της εικόνας, διαχωρισμό πλάνων του βίντεο κ.ά. Στο επόμενο στάδιο εξάγονται αριθμητικά χαρακτηριστικά από το περιεχόμενο, τα οποία χρησιμοποιούνται για την ομαδοποίηση και την



Σχήμα 1.1: Σχηματικό διάγραμμα ανάλυσης περιεχομένου με βοήθεια γνώσης.

επιλογή τμημάτων ενδιαφέροντος. Σε αυτό το στάδιο μπορούν να εφαρμοστούν μετασχηματισμοί του σήματος, όπως π.χ. μετασχηματισμός Fourier ή μετασχηματισμός κυματιδίου και έπειτα να εξαχθούν διανύσματα χαρακτηριστικών, όπως είναι οι οπτικοί ή οι ακουστικοί περιγραφείς. Τα αριθμητικά αυτά χαρακτηριστικά αναλύονται με μεθοδολογίες μηχανικής μάθησης και αναγνώρισης προτύπων και με τη συνδρομή της πολυμεσικής γνώσης που είναι αποθηκευμένη σε μορφή οντολογίας αντιστοιχίζονται σε ένα αρχικό σύνολο από σημασιολογικές οντότητες. Στη συνέχεια, η αξιοποίηση της τυπικής σημασιολογίας και της συλλογιστικής βελτιώνει την αρχική σημασιολογική ανάλυση, διορθώνοντας εσφαλμένες εκτιμήσεις, επιλύοντας τυχόν αντικρουόμενα δεδομένα, εξάγοντας νέα έμμεση γνώση και παρέχοντας εντέλει μία συνολικά συνεπή, σημασιολογική περιγραφή του πολυμεσικού υλικού. Με βάση τη σημασιολογική αναπαράσταση του πολυμεσικού περιεχομένου, ο τελικός στόχος είναι η δεικτοδότηση και ευφυής αναζήτηση πολυμεσικού περιεχομένου, δηλαδή η ανάπτυξη μηχανισμών για σημασιολογική αναζήτηση, ανάκτηση, κατηγοριοποίηση και πλοήγηση του περιεχομένου.

Το αντικείμενο της παρούσας διατριβής κινείται μέσα στο παραπάνω περιγραφέν πλαίσιο και πιο συγκεκριμένα ερευνάται η αξιοποίηση της γνώσης στη διαδικασία της κατάτμησης εικόνων και βίντεο και στον εντοπισμό και αναγνώριση εννοιολογικών οντοτήτων μέσα στο περιεχόμενο. Δίνεται έμφαση στην επιλογή της αναπαράστασης της γνώσης ώστε να εξυπηρετεί το διττό της ρόλο ως ενσωματωμένη διαδικασία σε αλγόριθμους ανάλυσης και ως λεξιλόγιο για τη σημασιολογική περιγραφή του περιεχομένου.

1.2 Διάρθρωση Κειμένου

Τα κεφάλαια που συνιστούν τη διατριβή αυτή αντιστοιχούν σε διαφορετικά στάδια της ανάλυσης περιεχομένου με βάση τη γνώση. Πιο συγκεκριμένα, στο κεφάλαιο 2 παρουσιάζουμε την απαραίτητη θεωρία σχετικά με δομές και γλώσσες αναπαράστασης γνώσης. Σε αυτές περιλαμβάνεται η ασαφής συνολοθεωρία που προσφέρεται ως μέσο αναπαράστασης αβεβαιότητας, η οποία είναι έμφυτη στα προβλήματα της όρασης υπολογιστών και ανάλυσης εικόνων και βίντεο. Τα ασαφή σύνολα δίνουν τη δυνατότητα χρήσης γλωσσικών όρων και κανόνων τους οποίους αξιοποιούμε για την αναπαράσταση αλλά και επεξεργασία των εννοιών που εξάγονται από το περιεχόμενο. Μία ευρέως διαδεδομένη μορφή αποθήκευσης δεδομένων αποτελούν οι γράφοι, και

ειδικά οι σχεσιακοί γράφοι χαρακτηριστικών. Η θεωρία γράφων μελετά τις σχέσεις αντικειμένων τα οποία έχουν κάποια συνάφεια μεταξύ τους ή συνδέονται με κάποιον ορισμένο τρόπο. Είναι προφανές ότι από παρόμοιες αρχές διέπονται τα αντικείμενα και οι οντότητες που υπάρχουν σε μία εικόνα ή μία ακολουθία βίντεο και αυτή ακριβώς την αίσθηση θεμελιώνουμε και χρησιμοποιούμε. Το κεφάλαιο 2 ολοκληρώνεται με την παρουσίαση των θεμελιωδών στοιχείων ενός οντολογικού μοντέλου γνώσης και την περιγραφή ορισμένων διαδεδομένων γλωσσών αναπαράστασης, οι οποίες χρησιμοποιούνται για την κατασκευή, την κωδικοποίηση και τη χρήση μίας οντολογίας.

Στο κεφάλαιο 3 περιγράφουμε τα στάδια ανάλυσης εικόνων με χρήση οντολογιών, ξεκινώντας από την προεπεξεργασία μίας εικόνας ή μίας ακολουθίας βίντεο, την εξαγωγή χαρακτηριστικών αυτής και καταλήγοντας στην ταξινόμηση με βάση τις έννοιες που εξήχθησαν από περιοχές της εικόνας. Σε αυτό το κεφάλαιο περιγράφεται η επικοινωνία του συστήματος ανάλυσης με τις οντολογίες πολυμέσων. Εν προκειμένω, κατά τη διαδικασία ανάλυσης πολυμεσικού περιεχομένου και της εξαγωγής της σημασιολογικής περιγραφής του, είναι αναγκαίο να ανακτηθεί η πρότερη γνώση από τις οντολογίες (οι οποίες χρησιμεύουν ως βάσεις γνώσεις). Πιο συγκεκριμένα, οι έννοιες της οντολογίας θεματικού πεδίου είναι συνδεδεμένες με πρότυπα στιγμιότυπα οπτικών περιγραφών μέσω της οντολογίας οπτικών περιγραφών η οποία μοντελοποιεί έννοιες και ιδιότητες που περιγράφουν τα οπτικά χαρακτηριστικά ενός αντικειμένου (σχήμα, χρώμα, υφή, κίνηση, κτλ.). Η ανάκτηση αυτών των στιγμιότυπων και το ταίριασμα των οπτικών περιγραφών τους με τους αντίστοιχους που εξάγονται από την υπό ανάλυση εικόνα ή βίντεο, αποτελεί ένα σημαντικό κομμάτι της αλληλεπίδρασης των αλγορίθμων ανάλυσης με τις τεχνολογίες γνώσεις.

Στη συνέχεια, στο κεφάλαιο 4 περιγράφουμε τη σημασιολογική επέκταση αλγορίθμων κατάτμησης, με σκοπό τον εντοπισμό και την αναγνώριση αντικειμένων μέσα σε εικόνες και ακολουθίες βίντεο. Πρόκειται για μία προσέγγιση η οποία εμπλουτίζει τεχνικές κατάτμησης επέκτασης περιοχών, με σημασιολογικά χαρακτηριστικά, προκειμένου να επιλυθούν προβλήματα υπερκατάτμησης ή υποκατάτμησης που συχνά παρατηρούνται και οφείλονται στην αδυναμία περιγραφής σημασιολογικών οντοτήτων με οπτικά μόνο χαρακτηριστικά. Πιο συγκεκριμένα, ο κύριος στόχος της σημασιολογικής επέκτασης περιοχών είναι να επιτύχει τόσο καλύτερη κατάτμηση της εικόνας, όσο και να χαρακτηρίσει και να κατατάξει τις επιμέρους περιοχές που συνθέτουν την εικόνα. Επιπρόσθετα, στο κεφάλαιο 4 παρουσιάζουμε μία καινοτόμο τεχνική για τη χωροχρονική κατάτμηση ακολουθιών βίντεο, χρησιμοποιώντας οπτικά και σημασιολογικά κριτήρια. Το αποτέλεσμα είναι η σημασιολογική χωροχρονική κατάτμηση, δηλαδή ο εντοπισμός, η παρακολούθηση και η αναγνώριση αντικειμένων ενδιαφέροντος μέσα από ακολουθίες βίντεο μικρής ή μεγάλης διάρκειας.

Στο κεφάλαιο 5 εξετάζουμε ένα τυπικό, δομημένο πλαίσιο γνώσης καθώς και την απαραίτητη μεθοδολογία που πρέπει να ακολουθηθεί για να επιτευχθεί η επιδιωκόμενη σημασιολογική κατάτμηση εικόνων και αναγνώρισης αντικειμένων. Ως γλώσσα αναπαράστασης γνώσης επιλέγουμε τις περιγραφικές λογικές και επιπλέον χρησιμοποιούμε τις ασαφείς επεκτάσεις τους για να μπορέσουμε να εντάξουμε στο σύστημά μας το βαθμό ασάφειας που υπεισέρχεται αναπόφευκτα στο στάδιο της ανάλυσης της εικόνας. Εφαρμόζουμε υπηρεσίες ασαφούς συλλογιστικής απευθείας στο πρόβλημα της κατάτμησης εικόνων, ως μέθοδο υπολογισμού σημασιολογικής απόστασης δύο περιοχών. Θέλοντας να αποφύγουμε το διαχωρισμό του προβλήματος της κατάτμησης και της αναγνώρισης, σε προβλήματα “χαμηλού” και “υψηλού” επιπέδου, εισάγουμε την έννοια της *οπτικής λογικής* για την ενιαία αντιμετώπιση αυτών των προβλημάτων.

Κεφάλαιο 1. Εισαγωγή

Τέλος, στο κεφάλαιο 6 συνοψίζουμε τα βασικά συμπεράσματα που παρουσιάστηκαν στα προηγούμενα κεφάλαια της διατριβής. Αναδεικνύουμε τη συνεισφορά μας και προτείνουμε κατευθύνσεις για περαιτέρω ερευνητικές επεκτάσεις.

□

Κεφάλαιο 2

Αναπαράσταση Γνώσης

2.1 Εισαγωγή

Ένα από τα παλαιότερα ζητήματα που έχουν απασχολήσει την επιστήμη των υπολογιστών, είναι η μελέτη της φυσικής γνώσης που βιωματικά αποκτά ένας άνθρωπος καθώς και ο τρόπος καταγραφής, αναπαράστασης και ενσωμάτωσης της σε μία μηχανή ή ένα υπολογιστικό σύστημα. Η διαδικασία της μάθησης και της συλλογιστικής, η οποία έχει αποτελέσει φιλοσοφικό θέμα από την αρχαιότητα, αποτελεί ένα ερευνητικό αντικείμενο με ρίζες σε πολλά επιστημονικά πεδία, όπως η ψυχολογία, η φιλοσοφία, η νευρολογία και η επιστήμη των υπολογιστών. Όσον αφορά το τελευταίο, έχει ενταχθεί στον τομέα της τεχνητής νοημοσύνης, η οποία μελετά τη σχεδίαση ευφυών υπολογιστικών συστημάτων που επιδεικνύουν χαρακτηριστικά που σχετίζονται με τη νοημοσύνη στην ανθρώπινη συμπεριφορά, όπως π.χ. μάθηση, προσαρμοστικότητα, εξαγωγή συμπερασμάτων, κατανόηση από τα συμφραζόμενα κ.ά. [103].

Η χρήση τεχνικών τεχνητής νοημοσύνης είναι ιδιαίτερα διαδεδομένη σε διαδικασίες υψηλού επιπέδου της όρασης υπολογιστών, όπως είναι η κατάτμηση εικόνων και βίντεο, η αναγνώριση αντικειμένων και καταστάσεων, η 3Δ ανακατασκευή σκηνής κ.ά. Για την επίλυση προβλημάτων αυτού του είδους απαιτείται καταρχάς μία αποδοτική μορφή αναπαράστασης (πρότερης) γνώσης και μία στρατηγική αξιοποίησης της γνώσης αυτής στο πλαίσιο ενός αλγορίθμου. Η αναπαράσταση γνώσης οφείλει (α) να είναι τυπική (να έχει αυστηρή σύνταξη και μονοσήμαντη ερμηνεία), (β) να είναι εκφραστική όσον αφορά τα σημασιολογικά χαρακτηριστικά της και (γ) να είναι όσο το δυνατόν περισσότερο αποδοτική έχουσα μικρό βαθμό πολυπλοκότητας. Τα τρία αυτά χαρακτηριστικά είναι πολλές φορές αντικρουόμενα, οπότε εξαρτάται από το σχεδιαστή η επιλογή της κατάλληλης αναπαράστασης γνώσης για το εκάστοτε πρόβλημα.

Στο παρόν κεφάλαιο παρουσιάζουμε την απαραίτητη θεωρία σχετικά με τις δομές και τις γλώσσες αναπαράστασης γνώσης που χρησιμοποιήσαμε στη διατριβή. Σε αυτές περιλαμβάνεται η ασαφής συνολοθεωρία που προσφέρεται ως μέσο αναπαράστασης αβεβαιότητας, η οποία είναι έμφυτη στα προβλήματα της όρασης υπολογιστών και ανάλυσης εικόνων και βίντεο. Τα ασαφή σύνολα δίνουν τη δυνατότητα χρήσης γλωσσικών όρων και κανόνων τους οποίους αξιοποιούμε για την αναπαράσταση αλλά και επεξεργασία των εννοιών που εξάγονται από το περιεχόμενο. Μία ευρέως διαδεδομένη μορφή αποθήκευσης δεδομένων αποτελούν οι γράφοι, και ειδικά οι σχεσιακοί γράφοι χαρακτηριστικών. Η θεωρία γράφων μελετά τις σχέσεις αντικειμένων τα οποία έχουν κάποια συνάφεια μεταξύ τους ή συνδέονται με κάποιον ορισμένο τρόπο. Είναι προφανές ότι από παρόμοιες αρχές διέπονται τα αντικείμενα και οι οντότητες που υπάρχουν

σε μία εικόνα ή μία ακολουθία βίντεο και αυτή ακριβώς την αίσθηση θεμελιώνουμε και χρησιμοποιούμε. Το κεφάλαιο κλείνει με την παρουσίαση του μαθηματικού μοντέλου οντολογικής γνώσης, που αποτελεί ένα σύγχρονο φορμαλισμό αποθήκευσης και αναπαράστασης γνώσης, και μία σύντομη επισκόπηση ορισμένων ευρέως χρησιμοποιούμενων γλωσσών αναπαράστασης οντολογιών.

2.2 Ασαφή Σύνολα

Η αβεβαιότητα αφθονεί στα προβλήματα της όρασης υπολογιστών και της ανάλυσης εικόνων και βίντεο και ως εκ τούτου η διαχείρισή της είναι μία πολύ σημαντική παράμετρος όταν σχεδιάζονται αλγόριθμοι και συστήματα για την επίλυση αυτών των προβλημάτων. Η αβεβαιότητα υπεισέρχεται σε διάφορα στάδια ενός συστήματος όρασης υπολογιστών: Προσθετικός θόρυβος κατά τη διάρκεια της αποτύπωσης της εικόνας, ανακρίβεια σε υπολογισμούς, ασάφεια στον ορισμό των κλάσεων και στην ερμηνεία εννοιών, ασθενώς ορισμένα (ill-posed) προβλήματα, κ.ά. Πολύ συχνά χρειάζεται να συνδυαστούν ξεχωριστές μετρήσεις από διαφορετικές πηγές, να υπολογιστούν ποικίλα χαρακτηριστικά και να εφαρμοστούν συμπληρωματικές ή παραπληρωματικές τεχνικές: μία πρακτική η οποία δημιουργεί νέα ερωτήματα όσον αφορά το πώς θα επεξεργαστεί η πλεονάζουσα πληροφορία και πώς θα επιλυθούν τυχόν αντιθέσεις και ασυνέπειες.

Η θεωρία των ασαφών συνόλων (ή αλλιώς ασαφής συνολοθεωρία) προσφέρεται ως μέσον αναπαράστασης αυτού του είδους αβεβαιότητας στη διατύπωση αλγορίθμων, καθώς μπορεί να διαχειριστεί ευέλικτα και με μεγάλη επιτυχία την αναπαράσταση ανακριβών μετρήσεων προσφέροντας μία λογική αναπαράσταση (σε φυσική γλώσσα) των αβέβαιων εννοιών με τη χρήση γλωσσικών κανόνων. Για παράδειγμα, η έννοια “καλός καιρός” με την παραδοσιακή λογική περιγράφεται με ένα σύνολο ποσοτικών μετρήσεων της θερμοκρασίας, υγρασίας, ηλιοφάνειας κ.ά., ενώ με χρήση ασαφούς λογικής θα μπορούσε να περιγραφεί ως “ζέστη με λιακάδα”. Η χρήση λεκτικών όρων προσφέρει μία προοδευτική μετάβαση από τη μία έννοια σε μία άλλη, κάτι που ανταποκρίνεται στην πραγματικότητα, καθώς τις περισσότερες φορές δεν είναι δυνατό να τραβηχτεί μία διαχωριστική γραμμή μεταξύ δύο εννοιών, όπως π.χ. “λιακάδα” και “συννεφιά”: ποσοτικά τι ποσοστό του ουρανού πρέπει να είναι καλυμμένο με σύννεφα για να θεωρηθεί ο καιρός συννεφιασμένος;

2.2.1 Ορισμοί και συμβολισμοί

Σε αυτή την ενότητα θα περιγράψουμε εν συντομία ορισμένες βασικές έννοιες της θεωρίας ασαφών συνόλων. Επιλέχθηκε να μη γίνει εκτεταμένη περιγραφή της θεωρίας, αλλά μία εισαγωγή σε κύριες έννοιες και συμβολισμούς, για την καλύτερη δυνατή κατανόηση των επόμενων κεφαλαίων της διατριβής. Για λεπτομερή περιγραφή της θεωρίας ασαφών συνόλων και ασαφούς λογικής ενδείκνυται το βιβλίο των Klir και Yuan [58], το οποίο αποτέλεσε τη βασική πηγή για τη συγκεκριμένη ενότητα. Τέλος για εφαρμογές των ασαφών συνόλων και συστημάτων στην όραση υπολογιστών και επεξεργασία εικόνας προτείνονται τα [15, 144].

Ορισμός 2.1 (Ασαφές Σύνολο) Ένα ασαφές σύνολο A ορίζεται ως προς το γενικό σύνολο (universal set) \mathbb{X} και αναπαριστάται με μία συνάρτηση συμμετοχής μ_A :

$$\mu_A: x \in \mathbb{X} \rightarrow \mu_A(x) \in [0, 1]. \quad (2.1)$$

Σε αντιδιαστολή με τα κλασικά σύνολα, όπου η χαρακτηριστική συνάρτηση I_A απεικονίζει τα στοιχεία $x \in \mathbb{X}$ στο δίτιμο σύνολο $\{0, 1\}$, η συνάρτηση συμμετοχής $\mu_A(x)$ τα απεικονίζει σε έναν πραγματικό αριθμό ο οποίος συμβολίζει το βαθμό συμμετοχής με τον οποίο το στοιχείο x ανήκει στο ασαφές σύνολο $A \subseteq \mathbb{X}$. Ας σημειωθεί ότι, χάριν απλούστευσης, συχνά για το συμβολισμό της συνάρτησης συμμετοχής $\mu_A(x)$ χρησιμοποιείται και ο $A(x)$, δηλαδή $A(x) \equiv \mu_A(x)$.

Ορισμός 2.2 (Στήριγμα Ασαφούς Συνόλου) Στήριγμα (*support*) ενός ασαφούς συνόλου A ως προς το γενικό σύνολο \mathbb{X} είναι το κλασικό σύνολο το οποίο περιέχει όλα τα στοιχεία του \mathbb{X} με μη μηδενικό βαθμό συμμετοχής στο A :

$$\text{supp}(A) = \{x \in \mathbb{X} | \mu_A(x) > 0\}. \quad (2.2)$$

Ένα ασαφές σύνολο A (με πεπερασμένο στήριγμα) είναι κατ' ουσίαν ένα σύνολο διατεταγμένων ζευγών, του στοιχείου $x \in \mathbb{X}$ και του βαθμού συμμετοχής του $\mu_A(x)$, δηλαδή $A = \{(\mu_A(x), x) | x \in \mathbb{X}\}$. Μία συνήθης και πρακτική αναπαράσταση ενός ασαφούς συνόλου A είναι αυτή του αθροίσματος:

$$A = \sum_{i=1}^n \mu_A(x_i)/x_i = \{\mu_A(x_1)/x_1 + \mu_A(x_2)/x_2 + \dots + \mu_A(x_n)/x_n\}, \quad (2.3)$$

όπου $i \in \mathbb{N}_n$ και $n = |\text{supp}(A)|$.

Ορισμός 2.3 (Ύψος Ασαφούς Συνόλου) Ύψος (*height*) ενός ασαφούς συνόλου A είναι το μέγιστο των βαθμών συμμετοχής στο \mathbb{X} μεταξύ όλων των στοιχείων του A :

$$h(A) = \sup_{x \in \mathbb{X}} \mu_A(x). \quad (2.4)$$

Ένα ασαφές σύνολο A λέγεται κανονικό αν $h(A) = 1$, ενώ λέγεται υποκανονικό αν $h(A) < 1$.

Ορισμός 2.4 (Πληθυκότητα) Η πληθυκότητα (*cardinality*) $|A|$ ενός ασαφούς συνόλου A ορίζεται το άθροισμα όλων των βαθμών συμμετοχής των στοιχείων του A :

$$|A| \equiv \text{card}(A) = \sum_{x \in \mathbb{X}} \mu_A(x). \quad (2.5)$$

Ορισμός 2.5 (Ισότητα Ασαφών Συνόλων) Δύο ασαφή σύνολα A και B λέγονται ίσα, $A = B$, όταν κάθε στοιχείο $x \in \mathbb{X}$ έχει ίδιο βαθμό συμμετοχής στα A και B :

$$A = B \Leftrightarrow \mu_A(x) = \mu_B(x), \quad \forall x \in \mathbb{X}. \quad (2.6)$$

Ορισμός 2.6 (Υποσύνολο) Για δύο ασαφή σύνολα A και B ορίζεται η σχέση του υποσυνόλου \subseteq . Το A είναι υποσύνολο του B (ή αλλιώς ότι το A περιέχεται στο B) αν:

$$A \subseteq B \Leftrightarrow \mu_A(x) \leq \mu_B(x), \quad \forall x \in \mathbb{X}. \quad (2.7)$$

Πολύ χρήσιμη είναι η έννοια της απόστασης δύο ασαφών συνόλων. Αν $\mathcal{F}(\mathbb{X})$ είναι το ασαφές δυναμοσύνολο (power set) του \mathbb{X} , τότε αυτό περιέχεται στο μοναδιαίο κύβο n -διάστασης $[0, 1]^n$. Σε αυτόν τον n -διάστατο χώρο μπορεί να ορισθεί η Minkowski απόσταση δύο ασαφών συνόλων A και B ως εξής:

$$d_p(A, B) = \left(\frac{1}{n} \sum_{i=1}^n |\mu_A(x_i) - \mu_B(x_i)|^p \right)^{1/p} \quad (2.8)$$

Συνήθεις επιλογές για την τιμή του p είναι $p = 1$ (Manhattan ή city-block απόσταση), $p = 2$ (Ευκλείδεια απόσταση) και $p \rightarrow \infty$ (Chebyshev απόσταση) η οποία ισοδυναμεί με τον τελεστή \max .

2.2.2 Ασαφείς Τελεστές

Για να μπορέσουμε να χρησιμοποιήσουμε τα ασαφή σύνολα ως μέσο αναπαράστασης αβεβαιότητας, είναι απαραίτητο να ορίσουμε τους τελεστές συμπληρώματος, τομής και ένωσης ενός ή παραπάνω ασαφών συνόλων. Γενικεύοντας τους αντίστοιχους τελεστές από την κλασική θεωρία συνόλων, διαπιστώνουμε ότι δεν υπάρχει μοναδικός ορισμός για αυτούς, τουναντίον για κάθε έναν υπάρχει μία οικογένεια συναρτήσεων η οποία ικανοποιεί ένα σύνολο αξιωμάτων. Για ένα ασαφές σύνολο A , το ασαφές συμπλήρωμά του \bar{A} ορίζεται ως $\bar{A} \equiv c(A): [0, 1] \rightarrow [0, 1]$. Το πρότυπο συμπλήρωμα είναι το $c(a) = 1 - a$.

Μία συνάρτηση t ονομάζεται τριγωνική νόρμα (t -νόρμα) και μπορεί να χαρακτηριστεί ως ασαφής τομή αν και μόνο αν πληροί τις παρακάτω βασικές ιδιότητες. Ο τελεστής ασαφούς τομής πραγματοποιείται από συναρτήσεις της μορφής $t: [0, 1] \times [0, 1] \rightarrow [0, 1]$, οπότε έχουμε $(A \cap B)(x) = t(A(x), B(x))$, $\forall x \in \mathbb{X}$. Ισχύει ότι η συνάρτηση t (και η u όπως θα δούμε λίγο παρακάτω) είναι ανεξάρτητη από το x , οπότε για απλότητα μπορούμε να αντικαταστήσουμε τις τιμές $A(x)$ και $B(x)$ με δύο τυχαίες τιμές $a, b \in [0, 1]$. Οπότε οι παρακάτω ιδιότητες είναι αναγκαίες και ικανές για τον ορισμό μίας ασαφούς τομής:

1. $t(a, 1) = a$ (ουδέτερο στοιχείο)
2. $t(a, 0) = 0$ (απορροφητικό στοιχείο)
3. $a \leq b \Rightarrow t(a, c) \leq t(b, c)$ (μονοτονικότητα)
4. $t(a, b) = t(b, a)$ (αντιμεταθετικότητα)
5. $t(a, t(b, c)) = t(t(a, b), c)$ (προσεταιριστικότητα)

Αντίστοιχα ορίζεται και η ασαφής ένωση $u: [0, 1] \times [0, 1] \rightarrow [0, 1]$ ως $(A \cup B)(x) = u(A(x), B(x))$, $\forall x \in \mathbb{X}$. Η u είναι t -κονόρμα ανν είναι αντιμεταθετική, προσεταιριστική, μη αύξουσα και για τα δύο ορίσματα, έχει το 0 για ουδέτερο και το 1 για απορροφητικό στοιχείο. Οι t -νόρμες και οι t -κονόρμες είναι *δυϊκές*, δηλαδή ικανοποιούν τις δύο παρακάτω σχέσεις:

$$c(u(a, b)) = t(c(a), c(b)) \quad (2.9)$$

$$c(t(a, b)) = u(c(a), c(b)) \quad (2.10)$$

Στον πίνακα 2.1 παρουσιάζονται ορισμένες ευρέως χρησιμοποιούμενες t -νόρμες και οι *δυϊκές* τους (ως προς το πρότυπο συμπλήρωμα) t -κονόρμες.

Πίνακας 2.1: Ασαφείς τελεστές: t -νόρμες και οι δυϊκές τους t -κονόρμες, ως προς το πρότυπο συμπλήρωμα.

Όνομα t -νόρμας	Ορισμός t -νόρμας	Όνομα t -κονόρμας	Ορισμός t -κονόρμας
Πρότυπη τομή	$\min(a, b)$	Πρότυπη ένωση	$\max(a, b)$
Αλγεβρικό γινόμενο	ab	Αλγεβρικό άθροισμα	$a + b - ab$
Lukasiewicz	$\max(0, a + b - 1)$	Φραγμένο άθροισμα	$\min(1, a + b)$
Γινόμενο Hamacher	$\frac{ab}{a + b - ab}$	Άθροισμα Hamacher	$\frac{a + b - 2ab}{1 - ab}$
Δραστική τομή	$\begin{cases} b, & \text{αν } a = 1 \\ a, & \text{αν } b = 1 \\ 0, & \text{αλλιού} \end{cases}$	Δραστική ένωση	$\begin{cases} b, & \text{αν } a = 0 \\ a, & \text{αν } b = 0 \\ 1, & \text{αλλιού} \end{cases}$

2.2.3 Ασαφείς Σχέσεις

Μία σχέση αναπαριστά την παρουσία ή την απουσία μίας συσχέτισης, αλληλεπίδρασης ή διασύνδεσης μεταξύ των στοιχείων δύο ή παραπάνω συνόλων. Αυτός ο ορισμός μπορεί να γενικευτεί ώστε να επιτρέπει την ύπαρξη βαθμού ισχύος της συσχέτισης ή της αλληλεπίδρασης μεταξύ των στοιχείων. Οι βαθμοί συσχέτισης μπορούν να αναπαρασταθούν ως βαθμοί συμμετοχής μίας ασαφούς σχέσης, κατ' αναλογία των βαθμών συμμετοχής ενός στοιχείου σε ένα ασαφές σύνολο. Πιο συγκεκριμένα, μία ασαφής σχέση R από το σύνολο A στο σύνολο B είναι ένα ασαφές σύνολο επί του $A \times B$. Το $R(a, b)$ αντιστοιχεί στο βαθμό που το a σχετίζεται με το b μέσω της σχέσης R .

Μερικές από τις βασικές ιδιότητες είναι οι ακόλουθες: Η αντίστροφη σχέση ορίζεται ως εξής: $R^{-1}(a, b) \equiv R(b, a)$. Η σύνθεση $\sup -t$ ορίζεται ως εξής: $[A \overset{t}{\circ} B](a, b) \equiv \sup(A(a, c), B(c, b))$. Η ταυτοτική σχέση, $I(a, b) = 1, a = b, I(a, b) = 0, a \neq b$, είναι το ουδέτερο στοιχείο της σύνθεσης: $R \circ I = I \circ R = R, \forall R$. Συχνά συναντιούνται οι ιδιότητες της ανακλαστικότητας ($I \subseteq R$), συμμετρικότητας ($R = R^{-1}$) και μεταβατικότητας $\sup -t$ ($R \overset{t}{\circ} R \subseteq R$). Επίσης, οι ιδιότητες της ισοδυναμίας (ανακλαστική, συμμετρική και μεταβατική), συμβατότητας (ανακλαστική και συμμετρική), διάταξης (αντισυμμετρική και μεταβατική) και ημιδιάταξης (μεταβατική και ανακλαστική).

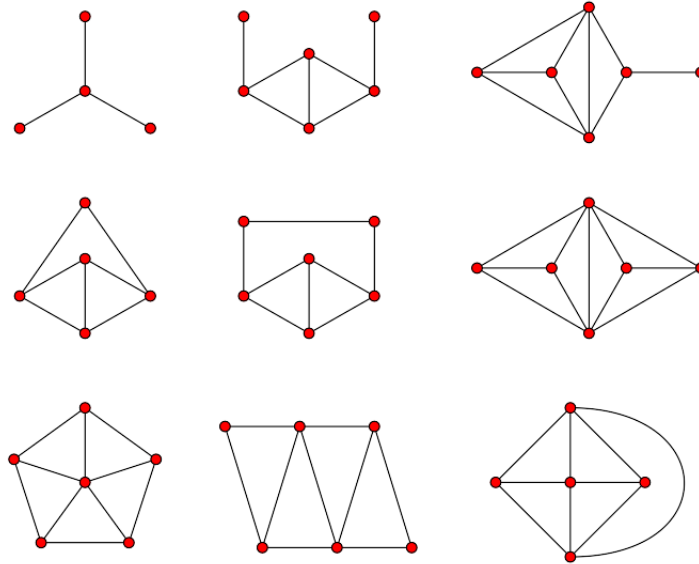
2.3 Θεωρία Γράφων

Η θεωρία γράφων (graph theory) αποτελεί ένα σημαντικό κομμάτι των διακριτών μαθηματικών, η οποία έχει βρει πάρα πολλές εφαρμογές στην επιστήμη των υπολογιστών, αλλά επίσης στη μηχανική, στις οικονομικές και ανθρωπιστικές επιστήμες. Η θεωρία γράφων [34, 142], εν γένει, μελετά τις σχέσεις αντικειμένων τα οποία έχουν κάποια συνάφεια μεταξύ τους ή συνδέονται με κάποιον ορισμένο τρόπο. Πολλά προβλήματα μπορούν να διατυπωθούν ως αφηρημένα προβλήματα της θεωρίας γράφων και της συνδυαστικής βελτιστοποίησης (combinatorial optimization), για τα οποία υπάρχουν κλασικές τεχνικές επίλυσης. Τυπικά:

Ορισμός 2.7 (Γράφος) Ένας γράφος $G \equiv (V, E)$ είναι ένα διατεταγμένο ζεύγος δύο ξένων μεταξύ τους συνόλων V και E . Τα στοιχεία του συνόλου $V \neq \emptyset$ ονομάζονται κόμβοι (ή κορυφές) του γράφου, ενώ τα στοιχεία του συνόλου $E \subseteq V \times V$ είναι

ένα υποσύνολο του συνόλου των μη διατεταγμένων ζευγών του V και ονομάζονται ακμές (ή πλευρές).

Ο συνήθης τρόπος αναπαράστασης ενός γράφου είναι να σχεδιάζεται μία κουκκίδα για κάθε κόμβο (vertex) και να ενώνονται δύο κουκκίδες με μία ευθεία αν οι αντίστοιχοι κόμβοι συνδέονται με μία ακμή (edge). Ο τρόπος που σχεδιάζονται οι κουκκίδες και οι γραμμές είναι αδιάφορος, σημαντική είναι μόνο η πληροφορία ποια ζεύγη κόμβων σχηματίζουν ακμές (σχήμα 2.1).



Σχήμα 2.1: Παραδείγματα επίπεδων γράφων.

Παρακάτω ορίζουμε μερικές βασικές έννοιες και την απαραίτητη μαθηματική σημειολογία περί γράφων, που θα χρησιμοποιήσουμε σε αυτό αλλά και σε επόμενα κεφάλαια της διατριβής.

Η πληθυκότητα του συνόλου των κόμβων V ονομάζεται *τάξη* του γράφου, ενώ η πληθυκότητα του συνόλου των ακμών E ονομάζεται *μέγεθος* του γράφου¹. Αν σε ένα γράφο G υπάρχει ακμή που ενώνει δύο κόμβους $v_a, v_b \in V$, τότε συμβολίζεται $e_{ab} \equiv (v_a, v_b) \in E$ και οι κόμβοι αυτοί λέγονται *γειτονικοί*. Η *γειτονιά* ενός κόμβου $v_a \in V$ ορίζεται ως το σύνολο των κόμβων που συνδέονται μαζί του:

$$N_a = \{v_b : e_{ab} \neq \emptyset\}. \quad (2.11)$$

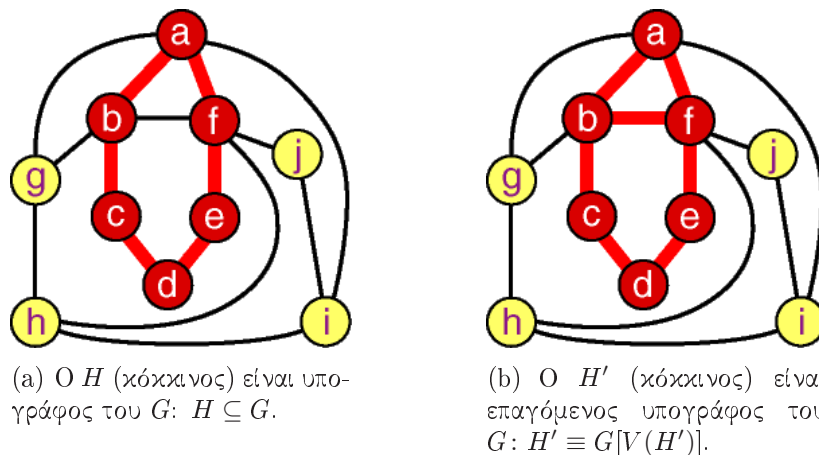
Ένας γράφος ονομάζεται *μη-κατευθυνόμενος* αν δύο κόμβοι $v_a, v_b \in V$ σχετίζονται αμοιβαία με ταυτόσημο τρόπο: $e_{ab} \equiv e_{ba}$, $\forall v_a, v_b \in V : e_{ab}, e_{ba} \in E$. Αν πάλι οι ακμές είναι προσανατολισμένες (ορίζονται από διατεταγμένα ζεύγη κόμβων), δηλαδή $e_{ba} \neq e_{ab}$, τότε ο γράφος ονομάζεται *κατευθυνόμενος*. Αν μία ακμή ενώνει μία κορυφή με τον εαυτό της, τότε λέγεται *βρόχος*, ενώ ακμές που έχουν τα ίδια άκρα λέγονται *πολλαπλές*. Οι πιο συνήθεις περιπτώσεις γράφων που θα μας απασχολήσουν είναι *απλοί*, δηλαδή γράφοι που δεν περιέχουν ούτε βρόχους ούτε πολλαπλές ακμές.

¹Θεωρούμε ότι το σύνολο των κόμβων πάνω στο οποίο ορίζεται ένας γράφος G συμβολίζεται ισοδύναμα ως $V \equiv V(G)$ και παρομοίως για το σύνολο των ακμών $E \equiv E(G)$. Μόνο όταν δεν είναι απόλυτα ξεκάθαρο από τα συμφραζόμενα σε ποιο γράφο αναφερόμαστε θα χρησιμοποιείται ο πιο αναλυτικός συμβολισμός $V(G)$ και $E(G)$.

Διαδρομή (path) λέγεται μια διάταξη κόμβων οι οποίοι ενώνονται με ακμές. Βαθμός (degree) d_a ενός κόμβου v_a σε ένα μη κατευθυνόμενο γράφο καλείται ο αριθμός των προσιπτουσών σε αυτόν ακμών, το σύνολο των οποίων ορίζεται ως:

$$E_a = \{e_{ab} : b \in N_a\} \subseteq E. \quad (2.12)$$

Πολύ χρήσιμη είναι η έννοια του υπογράφου (subgraph) $G' \equiv (V', E')$ ενός γράφου $G \equiv (V, E)$: είναι ο γράφος του οποίου όλες οι ακμές και οι κόμβοι περιέχονται στον G , δηλαδή $V' \subseteq V$ και $E' \subseteq E$ και συμβολίζεται $G' \subseteq G$. Δεδομένου ενός υποσυνόλου των κόμβων του γράφου G , $V' \subseteq V$, ο υπογράφος G' που προκύπτει από το V' και το $E' \equiv E \cap (V \times V')$ λέγεται επαγόμενος υπογράφος (induced subgraph) και συμβολίζεται $G' \equiv G[V']$. Στο σχήμα 2.2(a) απεικονίζονται δύο γράφοι: ο ένας είναι ο G με 10 κόμβους, $V(G) = \{a, b, c, d, e, f, g, h, i, j\}$ και ο άλλος ο H , με $V(H) = \{a, b, c, d, e, f\}$ (χρωματισμένος με κόκκινο). Είναι σαφής η σχέση $H \subseteq G$. Στο σχήμα 2.2(b) έχουμε τον ίδιο γράφο G ενώ για τον H' ισχύει $V(H') = V(H)$ και $E(H') = \{E(H), e_{bf}\}$ και είναι πλέον επαγόμενος υπογράφος του G .



Σχήμα 2.2: Σχέση υπογράφου και επαγόμενου υπογράφου.

Δύο γράφοι G και H λέγονται *ισομορφικοί* και συμβολίζονται $G \simeq H$, αν υπάρχει ένα προς ένα αντιστοιχία των κορυφών τους έτσι ώστε ένα ζεύγος κορυφών του G συνδέεται με μία ακμή ανν το αντίστοιχο ζεύγος κορυφών του H συνδέεται με μία ακμή. Συνεκτικός (connected) ονομάζεται ο γράφος για τον οποίο υπάρχει διαδρομή από κάθε κόμβο σε κάθε άλλο κόμβο. Ένας υπογράφος που περιέχει όλους του κόμβους και παραμένει συνεκτικός και ακυκλικός λέγεται *συνδετικό δένδρο* (spanning tree) για το οποίο ισχύουν οι παρακάτω ιδιότητες:

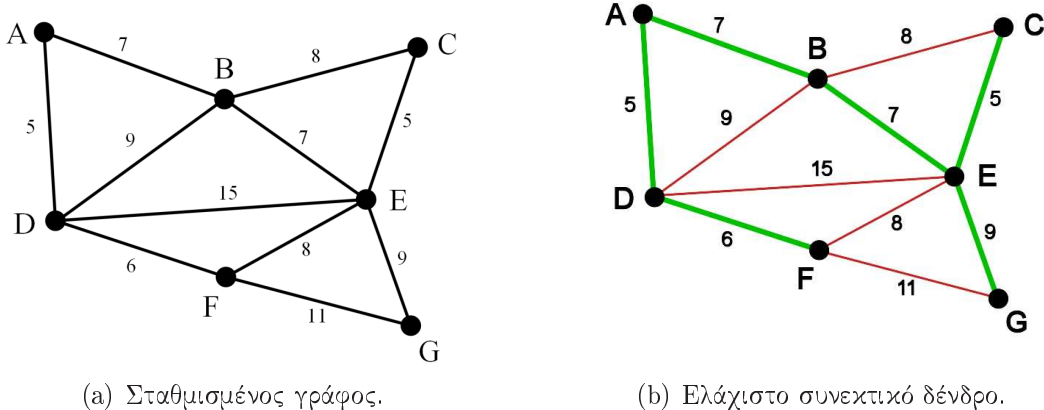
- Αν διαγραφεί οποιαδήποτε ακμή ο γράφος θα πάψει να είναι συνεκτικός.
- Δύο διαφορετικοί κόμβοι συνδέονται από μια και μόνο μια απλή διαδρομή.
- Ο γράφος περιέχει ακριβώς $|V| - 1$ ακμές.

Τέλος, μία σημαντική κατηγορία γράφων, την οποία θα χρησιμοποιήσουμε εκτεταμένα, είναι οι λεγόμενοι σταθμισμένοι γράφοι (weighted graphs), στους οποίους κάθε ακμή δε λογίζεται ισοδύναμη με όλες τις υπόλοιπες, αλλά σταθμίζεται διαφορετικά:

Ορισμός 2.8 (Σταθμισμένος Γράφος) Ένας γράφος G λέγεται σταθμισμένος αν στις ακμές του έχει οριστεί μία συνάρτηση βάρους (ή κόστους) $w: E \rightarrow \mathbb{R}$. Ανάλογα με την περίπτωση, η συνάρτηση βάρους w μπορεί να μοντελοποιεί απόσταση, χρόνο, ομοιότητα, και το πεδίο τιμών να μην είναι το \mathbb{R} αλλά το \mathbb{R}^+ ή το \mathbb{N} .

Οι σταθμισμένοι γράφοι έχουν ευρεία διάδοση σε εφαρμογές όπως η διαχείριση δικτύων, για αυτό και συχνά αποκαλούνται και δίκτυα. Στην περίπτωση των σταθμισμένων γράφων, το βάρος ενός μονοπατιού ισούται με το άθροισμα των βαρών των ακμών που συνιστούν το μονοπάτι. Τέλος, η συνάρτηση βάρους δεν υπόκειται στον περιορισμό της τριγωνικής ανισότητας $w(e_{ac}) \not\leq w(e_{ab}) + w(e_{bc})$.

Στο σχήμα 2.3(a) φαίνεται ένας σταθμισμένος γράφος 8 κόμβων και 17 ακμών, με τη συνάρτηση βάρους να έχει πεδίο τιμών το \mathbb{N} . Στον εν λόγω γράφο μπορούμε να υπολογίσουμε το ελάχιστο συνεκτικό δένδρο (minimum spanning tree), έστω T , με κριτήριο ελαχιστοποίησης το άθροισμα των βαρών των ακμών του δένδρου. Στο σχήμα 2.3(b) απεικονίζονται με πράσινο οι ακμές του ελάχιστου συνεκτικού δένδρου. Είναι προφανές ότι το T είναι συνεκτικό δένδρο του γράφου, αντιθέτως δεν είναι καθόλου προφανές ότι η επιλογή των ακμών ελαχιστοποιεί το άθροισμα των βαρών. Οι δύο πιο γνωστοί αλγόριθμοι (πολυωνυμικής πολυπλοκότητας) για εύρεση ελάχιστου συνεκτικού δένδρου είναι γνωστοί ως αλγόριθμος Kruskal και αλγόριθμος Prim [34].



(a) Σταθμισμένος γράφος.

(b) Ελάχιστο συνεκτικό δένδρο.

Σχήμα 2.3: Εξαγωγή ελάχιστου συνεκτικού δένδρου από σταθμισμένο γράφο.

2.3.1 Αναπαράσταση Εικόνας με Γράφο

Για την εφαρμογή ενός αλγορίθμου επεξεργασίας και ανάλυσης μίας εικόνας είναι απαραίτητη η κατάλληλη αναπαράστασή της σε μία δομή δεδομένων [121]. Η συνηθέστερη αναπαράσταση σε υπολογιστή είναι αυτή ενός πίνακα, όπου τα στοιχεία του πίνακα απεικονίζουν τη φωτεινότητα ή κάποια άλλη ιδιότητα του αντίστοιχου εικονοστοιχείου. Αυτή η απεικόνιση είναι απαραίτητη στη φάση της προεπεξεργασίας μίας εικόνας και αποτελεί την είσοδο σε αλγορίθμους επεξεργασίας χαμηλού επιπέδου, όπως φιλτραρίσματος, ανίχνευσης ακμών κ.ά.. Πολλές φορές θέλουμε να έχουμε μία υψηλότερου σημασιολογικού επιπέδου περιγραφή της εικόνας, όπου το βασικό στοιχείο της δομής δεν είναι ένα εικονοστοιχείο, αλλά μία ομάδα αυτών με κοινά χαρακτηριστικά όπως φωτεινότητα, χρώμα ή υφή, τα οποία πιθανώς αντιπροσωπεύουν ένα αντικείμενο ή μία περιοχή ενδιαφέροντος. Ουσιαστικά η περιγραφή της εικόνας

μεταφέρεται από το επίπεδο του εικονοστοιχείου στο επίπεδο των περιοχών, τα οποία συνήθως προκύπτουν από διαδικασίες συσταδοποίησης ή κατάτμησης.

Για μία πλήρη κατάτμηση της εικόνας I σε ένα πεπερασμένο σύνολο N περιοχών ή τμημάτων (regions/segments) $\{s_a\}$ ισχύει ότι $I = \bigcup_{a=1}^N s_a$ και $s_a \cap s_b = \emptyset, \forall a, b, a \neq b$.

Εφόσον μπορεί να γίνει η παραδοχή ότι μία εικόνα I μπορεί να περιγραφεί από τις επιμέρους περιοχές $\{s_a\}$ που την αποτελούν, ως ένα δομημένο σύνολο αντικειμένων, τότε είναι σαφής η δυνατότητα απεικόνισής της σε έναν (επίπεδο) γράφο, όπως αυτός ορίστηκε στην προηγούμενη ενότητα (ορισμός 2.7). Σε αυτήν την περίπτωση, το σύνολο V των κόμβων του γράφου G σχετίζεται με το σύνολο των περιοχών της εικόνας $\{s_a\}$, ενώ οι ακμές E στις τοπολογικές σχέσεις μεταξύ των περιοχών. Η τοπολογική σχέση της γειτνίασης δύο περιοχών ικανοποιείται αν οι περιοχές αυτές έχουν κοινό περίγραμμα.

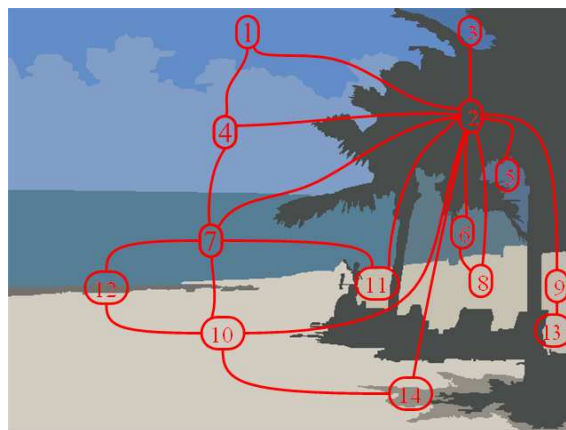
Στο σχήμα 2.4 φαίνεται πως γίνεται δυνατή η αντιστοίχιση των περιοχών μίας κατάτμησης S μίας εικόνας I με ένα γράφο. Το σύνολο των περιοχών της κατάτμησης της εικόνας $\{s_a\}$ έχουν αντιστοιχηθεί στους κόμβους $V = \{1, \dots, 14\}$, ενώ η γειτνίαση δύο περιοχών έχει αποτυπωθεί ως μία ακμή μεταξύ των αντίστοιχων κόμβων.



(a) Τυχαία εικόνα I .



(b) Μία κατάτμηση S της εικόνας I .



(c) Υπέρθηση του γράφου G στην κατάτμηση S

Σχήμα 2.4: Παράδειγμα αναπαράστασης κατάτμησης S μίας εικόνας με ένα σχεσιακό γράφο χαρακτηριστικών.

Σε αυτή την τοπογραφική αναπαράσταση των περιοχών της εικόνας είναι απαραίτητο να προσθέσουμε την πληροφορία της περιγραφής τόσο των περιοχών (με χρήση

οπτικών περιγραφών που θα παρουσιαστούν στο κεφάλαιο 3.4.2) όσο και πληροφορία για τη χωρική σχέση αυτών. Αυτό επιτυγχάνεται με την προσθήκη χαρακτηριστικών τόσο στους κόμβους όσο και στις ακμές του γράφου, και η δομή αυτή ονομάζεται *σχεσιακός γράφος χαρακτηριστικών* (attributed relational graph - ARG) [133, 36]. Οι κόμβοι και οι ακμές του ARG ενσωματώνουν τόσο συντακτική, όπως για παράδειγμα η αριθμητική-παραμετρική περιγραφή μίας περιοχής ή η χωρική διάταξη δύο περιοχών, όσο και σημασιολογική πληροφορία, όπως τι πιθανά αναπαριστά μία περιοχή, ή ποιο σύνθετο αντικείμενο αντιστοιχεί σε ένα υποσύνολο περιοχών (ισοδύναμα σε έναν υπογράφο του ARG της εικόνας).

Ορισμός 2.9 (Σχεσιακός Γράφος Χαρακτηριστικών) Ένας Σχεσιακός Γράφος Χαρακτηριστικών, ARG, ορίζεται από τις χωρικές οντότητες (περιοχές) s_a , με κάθε μία να αναπαριστάνεται από έναν κόμβο $v_a \in V$ και το διάνυσμα χαρακτηριστικών του κόμβου $\alpha(v_a)$, και τις δυαδικές χωρικές σχέσεις $e_{ab} \in E$ με το διάνυσμα χαρακτηριστικών της ακμής $w(e_{ab})$.

Η γενικότητα του παραπάνω ορισμού έγκειται στην επιλογή των χαρακτηριστικών $\alpha(\cdot)$ και $w(\cdot)$ των κόμβων και των ακμών αντίστοιχα. Στο κεφάλαιο 3 θα ασχοληθούμε με την επιλογή και την εξαγωγή των χαρακτηριστικών $\alpha(\cdot)$ και $w(\cdot)$.

Η αναπαράσταση μίας εικόνας με τον ARG έχει το πλεονέκτημα να συνδυάζει την τοπολογική περιγραφή που προσφέρει η δομή του γράφου, με την παραμετρική περιγραφή της εικόνας σε επίπεδο περιοχών, η οποία παρέχει αριθμητική αποτίμηση των χαρακτηριστικών, απαραίτητη για τη συντριπτική πλειοψηφία των τεχνικών ανάλυσης (συμπεριλαμβανομένου και αυτών της κατάτμησης και αναγνώρισης με τις οποίες θα ασχοληθούμε στα κεφάλαια 3 και 4).

2.4 Οντολογικό Μοντέλο Γνώσης

Οι *οντολογίες* (ontologies) [48, 14] αποτελούν ένα σύγχρονο και ευρέως διαδεδομένο φορμαλισμό (formalism) αποθήκευσης και αναπαράστασης γνώσης. Ο όρος προέρχεται από τη φιλοσοφία και στο εννοιολογικό πλαίσιο της επιστήμης των υπολογιστών ορίζεται ως ένας τυπικός και ρητός προσδιορισμός ενός λεξιλογίου που χρησιμοποιείται για να περιγράψει ένα θεματικό πεδίο με τις έννοιες, τις ιδιότητες και τις σχέσεις που το αποτελούν [48].

Εφόσον μία οντολογία αποτελείται από τις έννοιες, τις ιδιότητες και τις σχέσεις μεταξύ των εννοιών, μαθηματικά μπορεί να εκφραστεί ως εξής:

$$\mathcal{O} \equiv \{\mathcal{C}, \mathcal{R}\} \quad (2.13)$$

Στην εξίσωση (2.13), \mathcal{O} είναι η οντολογία, \mathcal{C} είναι το σύνολο των εννοιών που περιλαμβάνονται στην οντολογία και $\mathcal{R} = \{r_{pq}\}: \mathcal{C} \times \mathcal{C} \rightarrow \{0, 1\}$, $p, q \in \mathcal{C}$ είναι το σύνολο των δυαδικών σχέσεων του συνόλου των εννοιών \mathcal{C} . Γενικά, οι (σημασιολογικές) σχέσεις περιγράφουν ένα είδος συνάφειας μεταξύ δύο εννοιών· αν για ένα ζεύγος εννοιών $p, q \in \mathcal{C}$ ισχύει $r_{pq} = 1$ τότε λέμε ότι συνδέονται σημασιολογικά. Οι ιδιότητες των εννοιών μπορούν να αναπαρασταθούν και αυτές ως σχέσεις μεταξύ της έννοιας και της ιδιότητας όπως θα δούμε παρακάτω στο επόμενο κεφάλαιο στην ενότητα 3.3.

Για την κατασκευή, την κωδικοποίηση και τη χρήση μίας οντολογίας είναι απαραίτητη μία τυπική γλώσσα αναπαράστασης γνώσης. Για αυτόν το σκοπό δημιουργήθηκε

αρχικά η RDF (Resource Description Framework), η επέκτασή της RDFs (Resource Description Framework Schema) [141] και μετέπειτα η OWL (Web Ontology Language) [95] από τον οργανισμό προτυποποίησης του παγκόσμιου ιστού W3C. Στις παρακάτω παραγράφους περιγράφουμε τα χαρακτηριστικά των τριών αυτών γλωσσών, τη σύνταξή τους και τις σημασιολογικές δυνατότητες που προσφέρουν.

2.4.1 Η Γλώσσα Αναπαράστασης Γνώσης RDF

Η RDF είναι μια γλώσσα η οποία χρησιμοποιείται για την απλή περιγραφή πόρων (resources) του διαδικτύου. Η περιγραφή αυτή εστιάζεται κυρίως στην απόδοση μετα-πληροφορίας για τις οντότητες αυτές, όπως είναι η περιγραφή του τίτλου, του ονόματος, της ημερομηνίας δημιουργίας και άλλων χαρακτηριστικών κάποιου πόρου του διαδικτύου. Με την έννοια πόρος, αναφερόμαστε σε οποιαδήποτε οντότητα του παγκόσμιου ιστού, όπως είναι μια ιστοσελίδα, ένα τμήμα ή ένα σύνολο από ιστοσελίδες, ηλεκτρονικά αρχεία ή ακόμα και αντικείμενα τα οποία δεν είναι άμεσα διαθέσιμα στο διαδίκτυο, όπως είναι για παράδειγμα ένα βιβλίο.

Η RDF βασίζεται στην ιδέα ότι οι πόροι οι οποίοι πρέπει να περιγραφούν έχουν ιδιότητες με συγκεκριμένες τιμές. Έτσι λοιπόν μια μετα-πληροφορία για έναν πόρο αποτελείται από μια ιδιότητα και την τιμή που έχει ο πόρος για την ιδιότητα αυτή. Μια έκφραση για έναν πόρο ονομάζεται RDF πρόταση. Συνοψίζοντας λοιπόν, μια RDF πρόταση αποτελείται από μια τριπλέτα (triple) ενός υποκειμένου (subject), ιδιότητας (property) και αντικειμένου (object). Τη θέση του υποκειμένου καταλαμβάνει ο πόρος, τη θέση της ιδιότητας η ιδιότητα που του αποδίδουμε, ενώ τη θέση του αντικειμένου η τιμή που έχει ο πόρος αυτός για την ιδιότητα. Η τιμή αυτή μπορεί να είναι κάποιος άλλος πόρος ή κάποια τιμή δεδομένων. Συντακτικά οι προτάσεις αυτές δηλώνονται από μια διατεταγμένη τριπλέτα της μορφής, $\langle s, p, o \rangle$. όπου τα s, p και o αντιπροσωπεύουν το υποκείμενο, την ιδιότητα και το αντικείμενο αντίστοιχα, ενώ η τριπλέτα τερματίζεται με το σύμβολο της τελείας.

Ένα σύνολο από τριπλέτες RDF μπορούμε να το αντιληφθούμε και ως ένα γράφο. Σε αυτόν το γράφο τα αντικείμενα και τα υποκείμενα παίζουν το ρόλο των κόμβων ενώ οι ιδιότητες παίζουν το ρόλο των ακμών που τους συνδέουν. Όπως αναφέραμε και στην εισαγωγή, η RDF είναι μια γλώσσα αναπαράστασης γνώσης για το σημασιολογικό ιστό. Έτσι λοιπόν, το πρότυπο της RDF καθορίζει και μια σύνταξη η οποία έχει ως σκοπό οι RDF τριπλέτες να δομούνται με έναν τρόπο επεξεργάσιμο από υπολογιστικά συστήματα και εφαρμογές. Η σύνταξη αυτή δε θα μπορούσε να βασίζεται σε άλλο πρότυπο παρά μόνο στην XML. Η σύνταξη αυτή αναφέρεται ως RDF/XML σύνταξη [13]. Για λόγους γενικότητας η RDF χρησιμοποιεί αναφορές ομοιόμορφων αναγνωριστικών πόρων (Uniform Resource Identifier - URI references) για να προσδιορίσει τις οντότητες οι οποίες βρίσκονται στη θέση του υποκειμένου, της ιδιότητας και του αντικειμένου. Μία αναφορά URI (URIRef) αποτελείται από το URI και από ένα προαιρετικό αναγνωριστικό τμήματος (fragment identifier). Για παράδειγμα το URIRef <http://www.example.org/index.html#section2> αποτελείται από το URI <http://www.example.org/index.html> και από το αναγνωριστικό τμήματος Section2 το οποίο διακρίνουμε από το URI με τη χρήση του συμβόλου #.

Ας δούμε τώρα πως μπορούμε να ορίσουμε μετα-πληροφορία σε διάφορους πόρους χρησιμοποιώντας τη γλώσσα RDF. Έστω ότι θέλουμε να περιγράψουμε τη γνώση ότι η ιστοσελίδα <http://www.example.org/index.html> έχει δημιουργό τον John Smith. Στην πρόταση αυτή μπορούμε να διακρίνουμε ένα υποκείμενο, τη σελίδα, ένα

αντικείμενο τον John Smith και μια ιδιότητα, αυτή του δημιουργού της ιστοσελίδας. Αυτή η μετα-πληροφορία μπορεί να περιγραφεί σε μια τριπλέτα RDF ως εξής:

```
<http://.../index.html> <http://.../creator> <http://www.example.org/85740>,
όπου 85740 είναι ένα αντικείμενο που αντιπροσωπεύει τον John Smith. Παρατηρούμε
ότι η RDF χρησιμοποιεί URIs όχι μόνο για να προσδιορίσει τον πόρο τον οποίο
και περιγράφουμε αλλά και την ιδιότητα και το αντικείμενο. Όπως γίνεται αντιλη-
πτό ίσως είναι λίγο άβολο να γράφουμε συνέχεια ολόκληρα τα URI των οντοτήτων
που χρησιμοποιούμε. Έτσι λοιπόν πολλές φορές χρησιμοποιούνται συντομεύσεις σε
μορφή προθεμάτων (prefixes) τα οποία ονομάζονται QNames. Έτσι λοιπόν η παρα-
πάνω τριπλέτα μπορεί να γραφτεί ως: <ex:index.html><dc:creator><exstaff:85740>.
```

Όπως είναι φυσικό, αν επιθυμούμε, μπορούμε να προσδώσουμε επιπλέον μετα-πληροφορίες για τον αρχικό μας πόρο. Για παράδειγμα μπορούμε να πούμε ότι η ημερομηνία δημιουργίας της ιστοσελίδας ήταν η 19η Σεπτεμβρίου του 2006 ή ότι η γλώσσα στην οποία είναι γραμμένη είναι τα Αγγλικά. Οι πληροφορίες αυτές μπορούν να περιγραφούν με τις ακόλουθες τριπλέτες:

```
<ex:index.html> <exterm:creation-date> <"September 19, 2006">,
<ex:index.html> <exterm:language> <"English">.
```

Ας δούμε τώρα ποια μορφή έχει η γνώση μας αυτή σε RDF/XML σύνταξη. Η RDF/XML μορφή των παραπάνω προτάσεων είναι η ακόλουθη:

```
<?xml version="1.0" ?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:exterm="http://www.example.org/terms/">
  <rdf:Description rdf:about="http://www.example.org/index.html">
    <exterm:creation-date>September 19, 2006</exterm:creation-date>
    <exterm:language>English</exterm:language>
    <dc:creator rdf:resource="http://www.example.org/staffid/85740"/>
  </rdf:Description>
</rdf:RDF>
```

Από το προηγούμενο παράδειγμα παρατηρούμε ότι η χρήση πόρων στη θέση των αντικειμένων μας δίνει τη δυνατότητα να δημιουργήσουμε αλυσίδες από τριπλέτες RDF όπου το αντικείμενο της μιας τριπλέτας εμφανίζεται ως υποκείμενο της άλλης. Αυτό το χαρακτηριστικό είναι αρκετά χρήσιμο, καθώς οι τιμές που αποδίδουμε στην ιδιότητα ενός πόρου μπορούν να έχουν και αυτές με τη σειρά τους κάποια περιγραφή και άρα να αναπαριστούν μια περίπλοκη δομή. Για παράδειγμα στις προηγούμενες τριπλέτες αποδώσαμε ως τιμή της ιδιότητας ημερομηνία δημιουργίας το λεκτικό "September 19, 2006". Ίσως μια καλύτερη προσέγγιση να ήταν η ημερομηνία αυτή να αναπαρισταθεί με κάποιον πόρο, που θα ονομάζεται ημερομηνία και θα έχει τις ιδιότητες μήνας, ημέρα και έτος, οι οποίες στη συγκεκριμένη περίπτωση θα έχουν τις τιμές "September", "19" και "2006", αντιστοίχως. Έτσι λοιπόν μπορούμε να κάνουμε τις παρακάτω δηλώσεις:

```
<ex:index.html><exterm:creation-date><exdateid:85740>.
<exdateid:85740><exterm:month><September>.
<exdateid:85740><exterm:day><19>.
<exdateid:85740><exterm:year><2006>.
```

2.4.2 Η Γλώσσα Αναπαράστασης Γνώσης RDFs

Η RDF μας παρέχει τη δυνατότητα να δημιουργήσουμε απλές προτάσεις για τους πόρους τους οποίους θέλουμε να περιγράψουμε χρησιμοποιώντας ιδιότητες, τιμές και URIs για τον προσδιορισμό των συστατικών που συμμετέχουν σε μια πρόταση. Η RDF όμως δεν παρέχει δυνατότητα να ορίσουμε και να περιγράψουμε ένα επιπλέον λεξιλόγιο, το οποίο πιθανόν να επιθυμούμε να χρησιμοποιήσουμε στις εφαρμογές μας. Πιο συγκεκριμένα δεν έχουμε τη δυνατότητα να ορίσουμε τις κλάσεις (έννοιες) οι οποίες εμφανίζονται σε μια εφαρμογή. Επιπρόσθετα, είναι φυσικό να επιθυμούμε την περιγραφή των κλάσεων και των ιδιοτήτων μας δηλώνοντας σχέσεις υπαγωγής (υπο-εννοιών) ανάμεσά τους. Η γλώσσα η οποία παρέχει τη λειτουργικότητα αυτή είναι η RDFs [24]. Ουσιαστικά η RDFs παρέχει ένα επιπλέον λεξιλόγιο πάνω σε αυτό της RDF, περιλαμβάνοντας στοιχεία που προσδίδουν την επιπρόσθετη αυτή λειτουργικότητα.

Η δυνατότητα ορισμού κλάσεων υλοποιείται μέσω του στοιχείου `rdfs:Class` του λεξιλογίου της RDFs. Όπως είδαμε στην προηγούμενη ενότητα προκειμένου να δηλώσουμε ότι ένας πόρος είναι στιγμιότυπο μιας κλάσης χρησιμοποιήσαμε το στοιχείο `rdf:type`. Έτσι λοιπόν προκειμένου να δηλώσουμε μια νέα κλάση, έστω την `ex:MotorVehicle`, αυτό που έχουμε να κάνουμε είναι να δημιουργήσουμε την τριπλέτα `<ex:MotorVehicle><rdf:type><rdfs:Class>`, στην οποία η συντόμευση “`ex:`” αναπαριστά το URI `http://www.example.org/schemas/vehicles`. Εν συνεχεία μπορούμε να δηλώσουμε ότι ο πόρος `ex:Audi` είναι μέλος της κλάσης `ex:MotorVehicle` γράφοντας: `<ex:Audi><rdf:type><ex:MotorVehicle>`.

Με αυτόν τον τρόπο μπορούμε να δημιουργήσουμε το δικό μας λεξιλόγιο, δηλαδή να ορίσουμε το σύνολο των κλάσεων που εμφανίζονται σε μια εφαρμογή μας. Για παράδειγμα μπορούμε επιπρόσθετα να κάνουμε τις δηλώσεις:

`<ex:Van><rdf:type><rdfs:Class>`. και `<ex:Truck><rdf:type><rdfs:Class>`. Παρατηρήστε ότι προκειμένου να δηλώσουμε μια κλάση πρέπει να κάνουμε χρήση της μετα-μοντελοποίησης. Αυτό γιατί η κλάση `ex:Van` ορίζεται ως στιγμιότυπο της μετα-κλάσης `rdfs:Class`. Αλλά ακόμα περισσότερο εφόσον το στοιχείο `rdfs:Class` είναι προφανώς και αυτό μια κλάση πρέπει να ισχύει, και όντως σε κάθε RDFs αρχείο ισχύει, η τριπλέτα `<rdfs:Class><rdf:type><rdfs:Class>` [49].

Από τη στιγμή που έχουμε ορίσει τις κλάσεις, μπορεί να μας ενδιαφέρει να ορίσουμε σχέσεις υπαγωγής ανάμεσα σε αυτές. Στην RDFs αυτό επιτυγχάνεται με τη χρήση του στοιχείου `rdfs:subClassOf`. Έτσι λοιπόν μπορούμε να δημιουργήσουμε τις ακόλουθες προτάσεις:

`<ex:Van> <rdf:subClassOf> <ex:MotorVehicle>`.
`<ex:Truck> <rdf:subClassOf> <ex:MotorVehicle>`.

Η RDF/XML σύνταξη των παραπάνω ορισμών είναι η ακόλουθη:

```
<?xml version="1.0" ?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://example.org/schemas/vehicles">
  <rdfs:Class rdf:ID="MotorVehicle" />
  <rdfs:Class rdf:ID="Truck">
    <rdfs:subClassOf rdf:resource="#MotorVehicle" />
  </rdfs:Class>
  <rdfs:Class rdf:ID="Van">
```

```

<rdfs:subClassOf rdf:resource="#MotorVehicle"/>
</rdfs:Class>
</rdf:RDF>

```

Εκτός από τη δυνατότητα περιγραφής κλάσεων, μέσω σχέσεων υπαγωγής, η RDFs παρέχει επιπλέον τη δυνατότητα δημιουργίας περιγραφών για τις ιδιότητες που επιθυμούμε να χρησιμοποιήσουμε σε μια εφαρμογή μας. Πιο συγκεκριμένα μπορούμε να περιγράψουμε σχέσεις υπαγωγής ανάμεσα σε δυο ιδιότητες, αλλά επιπρόσθετα μπορούμε να ορίσουμε το πεδίο ορισμού και το πεδίο τιμών μιας σχέσης. Το πεδίο ορισμού μας δηλώνει το τι τύπου επιτρέπεται να είναι οι πόροι οι οποίοι βρίσκονται στη θέση του υποκειμένου μιας τριπλέτας ενώ το πεδίο τιμών μας λέει τι τύπου είναι οι πόροι οι οποίοι βρίσκονται στη θέση του αντικειμένου. Τα στοιχεία που παρέχονται από την RDFs για την υλοποίηση της λειτουργικότητας αυτής είναι τα στοιχεία `rdfs:subPropertyOf`, `rdfs:domain` και `rdfs:range`.

2.4.3 Η Γλώσσα Αναπαράστασης Γνώσης OWL

Η γλώσσα OWL [95] είναι μια γλώσσα αναπαράστασης γνώσης για το σημασιολογικό ιστό. Η γλώσσα OWL, στο μεγαλύτερο μέρος της, βασίζεται πάνω σε εκφραστικές περιγραφικές λογικές (ΠΛ) [11]. Το πρότυπο της OWL καθορίζει ουσιαστικά τρεις υπο-γλώσσες αυξανόμενης εκφραστικής δυνατότητας. Οι γλώσσες αυτές είναι οι ακόλουθες:

OWL Lite: Η γλώσσα αυτή απευθύνεται σε χρήστες οι οποίοι επιθυμούν να χρησιμοποιήσουν την OWL για την περιγραφή γνώσης σε εφαρμογές που δεν έχουν μεγάλες απαιτήσεις σε εκφραστικές δυνατότητες. Έτσι δίνεται η δυνατότητα ανάπτυξης εξειδικευμένων εργαλείων και μηχανισμών εξαγωγής συμπερασμάτων, τα οποία αναμένεται να λειτουργούν ταχύτερα από εργαλεία τα οποία υλοποιούν περισσότερο εκφραστικές γλώσσες. Η γλώσσα αυτή είναι ισοδύναμη με την περιγραφική λογική SHIF(D) [95].

OWL DL: Η γλώσσα αυτή δίνει τη μέγιστη εκφραστική δυνατότητα που προσφέρεται από τη γλώσσα OWL χωρίς όμως να χάνονται οι καλές υπολογιστικές ιδιότητές της. Αυτό σημαίνει ότι η γλώσσα αυτή, σε αντίθεση με την τελευταία υπο-γλώσσα της OWL, είναι αποφασίσιμη (decidable). Η γλώσσα αυτή είναι ισοδύναμη με την περιγραφική λογική SHION(D) [95].

OWL Full: Η γλώσσα αυτή προσφέρει το ίδιο λεξιλόγιο με τη γλώσσα OWL DL. Επιπρόσθετα όμως παρέχει τη συντακτική ελευθερία και τα χαρακτηριστικά της γλώσσας RDF και πιο συγκεκριμένα τη δυνατότητα μετα-μοντελοποίησης. Η γλώσσα αυτή είναι σαφώς μη-αποφασίσιμη (undecidable) καθώς δεν επιβάλλει περιορισμούς στη χρήση μεταβατικών ιδιοτήτων [83].

Αντίστοιχα με τις ΠΛ, η OWL περιλαμβάνει ένα αλφάβητο το οποίο αποτελείται από κλάσεις (classes), ιδιότητες (properties) και άτομα (individuals). Οι κλάσεις της OWL αποτελούν ένα ανάλογο των εννοιών των περιγραφικών λογικών. Έτσι λοιπόν διαισθητικά μια κλάση αναπαριστά ένα σύνολο από αντικείμενα τα οποία έχουν κοινά χαρακτηριστικά, όπως για παράδειγμα η κλάση των ανθρώπων ή η κλάση των αυτοκινήτων. Όπως οι ΠΛ, έτσι και η OWL, περιλαμβάνει τόσο ατομικές κλάσεις όσο και περιγραφές κλάσεων (class descriptions). Επιπρόσθετα, παρέχονται και αξιώματα κλάσεων (class axioms) όπως αξιώματα υπαγωγής, ισοδυναμίας αλλά και πολλά ακόμα. Οι ιδιότητες, που είναι το αντίστοιχο των ρόλων στις ΠΛ, αναπαριστούν δυαδικές σχέσεις, δηλαδή ζεύγη αντικειμένων. Επιπρόσθετα η OWL προσφέρει τη δυνατότητα ορισμού αξιωμάτων ιδιοτήτων (property axioms), όπως αξιώματα μετα-

βατικών ρόλων αλλά και άλλα αξιώματα τα οποία δεν εμφανίζονται άμεσα παρά μόνο έμμεσα στις ΠΛ. Τέλος η OWL προσφέρει τη δυνατότητα ορισμού ισχυρισμών, οι οποίοι στην περίπτωση της OWL ονομάζονται γεγονότα (facts). Τέλος κατά αντιστοιχία με τις ΠΛ η OWL προσφέρει έναν τυπικό τρόπο για την απόδοση ερμηνείας στα δομικά της συστατικά.

Όπως αναφέραμε και στην εισαγωγή εφόσον η OWL είναι μια γλώσσα αναπαράστασης γνώσης για το σημασιολογικό ιστό, εκτός από την αφηρημένη σύνταξη, πρέπει να διαθέτει και RDF/XML σύνταξη. Η σύνταξη αυτή μας δίνει τη δυνατότητα να διαμορφώσουμε τη γνώση μας σε μια μορφή προσπελάσιμη από εφαρμογές του παγκοσμίου ιστού. Εφόσον η OWL διαθέτει αρκετά περίπλοκους κατασκευαστές, η σύνταξη αυτή είναι πολλές φορές περίπλοκη.

2.4.4 Ασαφής Επέκταση της Γλώσσας OWL

Για την πλειοψηφία των εφαρμογών του σημασιολογικού ιστού είναι ζητούμενη η δυνατότητα διαχείρισης της ασάφειας και της αβεβαιότητας της υπάρχουσας πληροφορίας, καθώς είναι περισσότερο χρήσιμη από την απλή γνώση της ύπαρξης ή όχι μίας έννοιας ή μίας σχέσης εννοιών. Σε αυτό το πλαίσιο καθορίστηκε η ασαφής επέκταση της OWL, η f-OWL [127, 124], η οποία έχει ως κύρια διαφοροποίηση την ύπαρξη βαθμών συμμετοχής στους ισχυρισμούς (γεγονότα).

Η επέκταση της f-OWL απαιτεί και καθορισμό νέας σύνταξης, τόσο σε αφηρημένο επίπεδο όσο και σε επίπεδο σύνταξης RDF/XML, έτσι ώστε να είναι δυνατή η δήλωση του βαθμού συμμετοχής. Ορίζεται λοιπόν το στοιχείο “membership”, το οποίο αποτελείται από δύο επιμέρους στοιχεία, τα “ineqType” και “degree”. Το “ineqType” δηλώνει το είδος της ανισότητας (π.χ. \geq , \leq , ενώ το “degree” υποδηλώνει το βαθμό συμμετοχής του ατόμου στη συγκεκριμένη κλάση και έχει πεδίο τιμών το $[0, 1]$. Στην περίπτωση που παραλείπεται ο βαθμός, υπονοείται τύπος ανισότητας το “=” και βαθμός ίσος με 1. Όσον αφορά την RDF/XML σύνταξη, δίνεται παρακάτω ένα παράδειγμα περιγραφής ενός αυτοκινήτου, όπου έχουμε ασαφοποίηση μίας ιδιότητάς του (“luxuryCar”) και της σχέσης του με κάποιο άλλο αυτοκίνητο όσον αφορά την ταχύτητα (“faster”):

```
<owl:Description rdf:about="Porsche_911">
  <owl:type rdf:resource="luxuryCar" owl:ineqType=">"
    owlx:degree="0.8"/>
  <faster rdf:resource="Mini_Cooper" owl:ineqType=">"
    owlx:degree="0.9"/>
</owl:Description>
```

□

Κεφάλαιο 3

Εξαγωγή Σημασιολογικής Περιγραφής με Χρήση Τεχνολογιών Γνώσης

3.1 Εισαγωγή

Η ανάλυση εικόνας και βίντεο είναι ένα δύσκολο και κατά κανόνα ασθενώς ορισμένο πρόβλημα. Αυτό σημαίνει ότι δεδομένης μόνο της οπτικής πληροφορίας (δηλαδή το σήμα μιας εικόνας ή ενός βίντεο) δεν υφίσταται μοναδική γενική λύση στο πρόβλημα της ανάλυσης [105]. Προκειμένου να μετατραπεί το πρόβλημα σε καλά ορισμένο και να καταστεί δυνατή η επίλυσή του, απαιτείται ο περιορισμός του πλήθους των αποδεκτών λύσεων. Για το σκοπό αυτό έχει προταθεί στη σχετική βιβλιογραφία η εισαγωγή, στη διαδικασία της ανάλυσης, πρότερης (a priori) γνώσης. Η πρότερη γνώση για ένα θεματικό πεδίο τυπικά περιλαμβάνει τα σημαντικά αντικείμενα του πεδίου, τα χαρακτηριστικά τους, σχέσεις μεταξύ αυτών, κ.ά. Δοθείσης αυτής της γνώσης, είναι εφικτή η διάκριση και η κατηγοριοποίηση των σημασιολογικών οντοτήτων ενός θεματικού πεδίου, καθώς και η αναγνώριση των διαφορετικών απεικονίσεων που μπορούν αυτές να έχουν [30].

Οι διάφορες προσεγγίσεις που παρουσιάζονται στη βιβλιογραφία για το πρόβλημα της ανάλυσης εικόνας και βίντεο με χρήση πρότερης γνώσης διακρίνονται με κριτήριο τον τύπο της γνώσης που αξιοποιούν σε δυο γενικές κατηγορίες. Στην πρώτη κατηγορία ανήκουν οι προσεγγίσεις που κάνουν χρήση γνώσης που δεν είναι αυστηρά ορισμένη αλλά υπονοείται από ένα σύνολο παραδειγμάτων και εξάγεται από αυτό με επεξεργασία του και συχνά κάνοντας χρήση μεθόδων μηχανικής εκμάθησης [17]. Το κύριο πλεονέκτημα των μεθόδων αυτών είναι η δυνατότητα να αναπαραστήσουν συσχετίσεις μεταξύ οπτικών χαρακτηριστικών και εννοιών μέσω της διαδικασίας της εκπαίδευσης. Τεχνικές μηχανικής μάθησης όπως οι μηχανές διανυσμάτων στήριξης (support vector machines - SVMs) [138] είναι πολύ διαδεδομένες για την αναπαράσταση, δεικτοδότηση και ανάκτηση περιεχομένου [89, 5, 64]. Το βασικό μειονέκτημα αυτών των τεχνικών είναι ο αδιαφανής τρόπος αναπαράστασης της γνώσης, που καθιστούν πολύ δύσκολη έως αδύνατη την αξιοποίησή της σε άλλο θεματικό πεδίο, καθώς και την επέκτασή της με περισσότερες σημασιολογικές οντότητες.

Στη δεύτερη κατηγορία ανήκουν οι προσεγγίσεις που αξιοποιούν αυστηρά ορισμένη γνώση, με χρήση τεχνικών τυπικής αναπαράστασης. Ειδικότερα, στο πλαίσιο της ανάπτυξης του σημασιολογικού ιστού [14] υιοθετήθηκαν τυπικές γλώσσες αναπαράστασης γνώσης ώστε να είναι ευκολότερη η επέκταση και η επαναχρησιμοποίηση της γνώσης, καθώς και η εξαγωγή νέας γνώσης μέσω της συλλογιστικής. Στις εργα-

σίες των Maillot et al. [68, 67] προτείνεται μία μεθοδολογία για την κατασκευή και την αξιοποίηση βάσεων γνώσης σε προβλήματα όρασης υπολογιστών. Η προσέγγιση αυτή βασίζεται στο συσχετισμό εννοιών από οντολογίες θεματικού πεδίου με οπτικούς περιγραφείς, που σε συνδυασμό με κανόνες χρησιμεύουν στη σημασιολογική ανάλυση εικόνων. Οι περιγραφικές λογικές (ΠΛ), κωδικοποιημένες στη γλώσσα DAML+OIL, χρησιμοποιήθηκαν στην εργασία [107] για την κατασκευή ενός εποπτευόμενου συστήματος εκμάθησης, που χειρίζεται τις περιοχές της εικόνας ως στιγμιότυπα οπτικών περιγραφών και την αναγνώριση εννοιών ως πρόβλημα ταξινόμησης. Στο πρόβλημα της κατανόησης περιεχομένου, σημαντικές είναι οι εργασίες των Möller και Neumann, οι οποίοι παρουσίασαν τη δική τους προσέγγιση χρησιμοποιώντας τυπικές περιγραφικές γλώσσες και μοντελοποιώντας το πρόβλημα αυτό ως πρόβλημα συνεπαγωγής. Αρχικά, διερεύνησαν τις επιπλέον δυνατότητες που παρέχουν οι πρότυπες επεκτάσεις των περιγραφικών λογικών [80] και μετέπειτα επέκτειναν τις ΠΛ περιλαμβάνοντας χωρικές σχέσεις ώστε βασιζόμενοι στις υποθέσεις και τις χωρικές σχέσεις, να μπορούν να επεκτείνουν τα αποτελέσματα σύμφωνα με τη διαθέσιμη γνώση [90].

3.2 Επισκόπηση Μεθοδολογίας

Σε αυτό το κεφάλαιο περιγράφεται η επικοινωνία του συστήματος ανάλυσης με τις οντολογίες πολυμέσων, με στόχο την εξαγωγή σημασιολογικής περιγραφής των εικόνων. Εν προκειμένω, όταν πρόκειται να αναλυθεί μία νέα εικόνα και να εξαχθεί από αυτή η σημασιολογική περιγραφή της, είναι αναγκαίο να ανακτηθεί η πρότερη γνώση από τις οντολογίες (οι οποίες χρησιμεύουν ως βάση γνώσης). Στην ενότητα 3.3 περιγράφουμε τις οντολογίες που συνθέτουν τη γνώση που χρησιμοποιείται για την ανάλυση των εικόνων και την εξαγωγή σημασιολογικής περιγραφής. Πιο συγκεκριμένα, οι έννοιες που αποτελούν το λεξιλόγιο της σημασιολογικής περιγραφής ομαδοποιούνται ανά θεματική κατηγορία και ορίζονται στις οντολογίες θεματικού πεδίου (domain ontologies). Οι έννοιες αυτές συνδέονται με τα οπτικά χαρακτηριστικά μέσω πρότυπων στιγμιότυπων (prototype instances), τα οποία έχουν κατασκευαστεί εκ των προτέρων με επιβλεπόμενη διαδικασία εξαγωγής οπτικών περιγραφών και αντιστοίχιση αυτών με έννοιες από την οντολογία θεματικού πεδίου. Επιπλέον, οι έννοιες και οι ιδιότητες των οπτικών περιγραφών (όπως είναι οι περιγραφείς για το σχήμα, το χρώμα, την υφή, την κίνηση, κ.ά.) μοντελοποιούνται στην οντολογία οπτικών περιγραφών (visual descriptor ontology). Η ανάκτηση αυτών των πρότυπων και το ταίριασμα των οπτικών περιγραφών των με τους αντίστοιχους, που εξάγονται από την υπό ανάλυση εικόνα ή βίντεο, αποτελεί ένα σημαντικό κομμάτι της αλληλεπίδρασης των αλγορίθμων ανάλυσης με τις τεχνολογίες γνώσης.

Η προτεινόμενη μεθοδολογία ανάλυσης εικόνων με βάση τη γνώση, αρχικά κατατέμνει την εικόνα σε ομοιογενείς περιοχές. Επιλέγουμε να εργαστούμε με μικρές, ομοιογενείς, χωρικά συνεκτικές περιοχές οι οποίες παράλληλα έχουν μία δομή και χαρακτηριστικά, σε αντίθεση με τα μεμονωμένα εικονοστοιχεία. Για κάθε περιοχή, εξάγουμε ένα σύνολο οπτικών περιγραφών (κατά βάση περιγραφείς του προτύπου MPEG-7), οι οποίοι σχηματίζουν το διάλυσμα χαρακτηριστικών της υπό εξέταση περιοχής. Η προσέγγιση του προβλήματος σε αυτήν τη φάση είναι η ανάλυση με βάση περιοχές και όχι στο σύνολο της εικόνας ή με βάση σημεία ενδιαφέροντος, έτσι ώστε να δίνεται η δυνατότητα ακριβούς εντοπισμού των εννοιών στην εικόνα. Προκειμένου να αντιστοιχίσουμε αυτές τις περιοχές με κάποιες έννοιες από την οντολογία θεματικής περιοχής, ανακτούμε όλα τα πρότυπα στιγμιότυπα για κάθε έννοια, με εξα-

ντλητική αναζήτηση, και με βάση την οντολογία οπτικών περιγραφέων, διαβάζουμε τους οπτικούς περιγραφείς των εννοιών αυτών. Στη συνέχεια ταιριάζουμε τους οπτικούς περιγραφείς με τους αντίστοιχους εξαχθέντες περιγραφείς των περιοχών, με βάση συγκεκριμένες τεχνικές ταιριάσματος, και υπολογίζουμε το βαθμό κατά τον οποίο μία περιοχή μπορεί να χαρακτηριστεί από μία έννοια. Με αυτόν τον τρόπο, επιτυγχάνουμε το χαρακτηρισμό της εικόνας με σημασιολογικές έννοιες και μάλιστα με ακρίβεια όσον αφορά τη χωρική τους θέση.

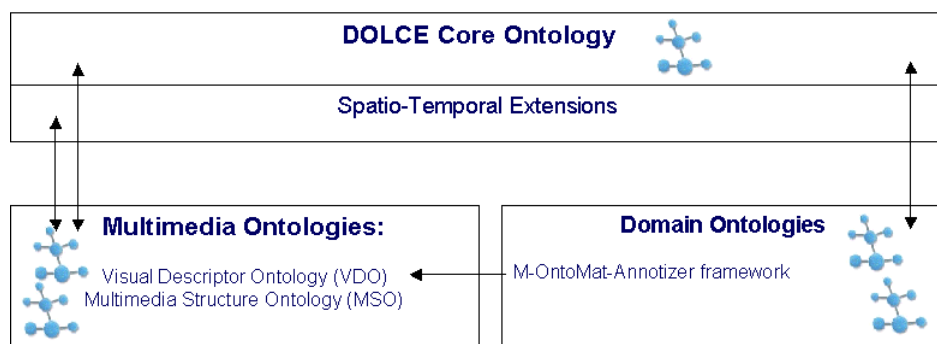
Τα παραγόμενα μεταδεδομένα, περιγράφουν σημασιολογικά το περιεχόμενο και μάλιστα με δομικό τρόπο, καθώς η περιγραφή γίνεται τόσο στο επίπεδο όλης της εικόνας, όσο και στο επίπεδο των αντικειμένων που περιέχονται σε αυτήν. Τα αντικείμενα εντοπίζονται μέσα στην εικόνα σε περιοχές που έχουν προκύψει από κατάτμηση και στις οποίες έχουμε αντιστοιχήσει έννοιες από το σχετικό θεματικό πεδίο μαζί με βαθμούς εμπιστοσύνης. Επίσης, εξάγονται και οι χωρικές σχέσεις (πάνω, κάτω, μέσα, κ.ά.) μεταξύ των αντικειμένων της εικόνας, έτσι κατ' αυτόν τον τρόπο τα μεταδεδομένα συνδυάζουν σημασιολογική και χωρική πληροφορία. Η σειριακοποίηση (serialization) των μεταδεδομένων γίνεται σε μορφή RDFs, ακολουθώντας το φορμαλισμό της οντολογίας δομής πολυμέσων (multimedia structure ontology), η οποία είναι μέρος της υποδομής γνώσης και θα περιγράψουμε στην ενότητα 3.3.

Η συνολική προσέγγιση μας στο πρόβλημα της ανάλυσης εικόνας για την εξαγωγή σημασιολογικών μεταδεδομένων με χρήση τεχνολογιών γνώσης αποτελεί ένα γενικό πλαίσιο, κλιμακούμενο (scalable) όσον αφορά τον αριθμό των εννοιών και των θεματικών πεδίων, τα οποία μπορεί να αναλύσει και να διαχειριστεί, καθώς η υποδομή των οντολογιών προσφέρει τη δυνατότητα επέκτασης με χρήση νέων οντολογιών θεματικών πεδίων.

3.3 Οντολογική Υποδομή Γνώσης

Σκοπός μας είναι η γνώση περιοχών να συνδυάζεται με τα χαμηλού επιπέδου οπτικά χαρακτηριστικά των αντικειμένων και τις χωρικές περιγραφές σχηματίζοντας έτσι ένα πλαίσιο ανάλυσης πολυμέσων υποβοηθούμενο από οντολογίες. Η γενική αρχιτεκτονική διασύνδεσης των οντολογιών, που απεικονίζεται στο σχήμα 3.1, αποτελείται από μια βάση γνώσης, συμπεριλαμβανομένου του σχήματος της οντολογίας, των αντίστοιχων στιγμιότυπων της θεματικής περιοχής και των κανόνων. Η κατασκευή της οντολογίας αφήνεται σε ειδικούς της οντολογικής ή πολυμεσικής μηχανικής, ενώ τα αποτελέσματα μπορούν να χρησιμοποιηθούν στη σημασιολογική δεικτοδότηση και ανάκτηση εικόνας και βίντεο, το σημασιολογικό σχολιασμό, κ.ά. Σύμφωνα με αυτήν την προσέγγιση, η διαδικασία ανάλυσης πολυμέσων εξαρτάται κατά ένα μεγάλο μέρος από τη βάση γνώσεων του συστήματος και κατά συνέπεια η μέθοδος μπορεί να εφαρμοστεί εύκολα σε διαφορετικές θεματικές περιοχές υπό τον όρο ότι η βάση γνώσεων είναι εμπλουτισμένη με τις αντίστοιχες πληροφορίες της θεματικής περιοχής. Ο εμπλουτισμός της βάσης γνώσης με χωροχρονικές σχέσεις αντικειμένων και ορισμούς γεγονότων παρέχει τα μέσα για επιπρόσθετα - εξαρτώμενα από το θεματικό πεδίο - εξαγωγή σημασιολογίας σε επίπεδο γεγονότων πάνω από τη σημασιολογία σε επιπέδων αντικειμένου (π.χ. ένα αυτοκίνητο που βγαίνει από το δρόμο).

Με την οικοδόμηση αυτού του ενοποιημένου μοντέλου που περιλαμβάνει όλες τις πτυχές για τη σημασιολογική ανάλυση πολυμέσων, όλα τα μέρη του αντιμετωπίζονται ως οντολογικές έννοιες. Συνεπώς, διαφορετικοί αλγόριθμοι ανάλυσης πολυμέσων μπορούν να εξεταστούν με απλό καθορισμό των αντίστοιχων στιγμιότυπων στην



Σχήμα 3.1: Υποδομή και διασύνδεση οντολογιών.

οντολογία και την ανάπτυξη κατάλληλων διεπαφών. Επιπλέον, νέες σημασιολογικές έννοιες μπορούν να καθοριστούν και να εξαχθούν αυτόματα με την περιγραφή των ιδιοτήτων τους, τα αντίστοιχα χαμηλού επιπέδου χαρακτηριστικά στην οντολογία και με τον εμπλουτισμό της βάσης γνώσης με τα κατάλληλα στιγμιότυπα. Κοινά εργαλεία οντολογιών μπορούν να χρησιμοποιηθούν για να διορθωθεί και να γίνει διαχείριση της προτεινόμενης οντολογικής υποδομής. Το σχήμα 3.1 απεικονίζει τη συνολική διασύνδεση των οντολογιών. Η οντολογική υποδομή αποτελείται από τέσσερις κατά βάση οντολογίες: (α) την οντολογία πυρήνα (core ontology), (β) την οντολογία πολυμέσων (multimedia ontology), (γ) τις οντολογίες θεματικών περιοχών (domain ontologies) και (δ) την οντολογία οπτικών περιγραφών (visual descriptor ontology). Στις παρακάτω παραγράφους περιγράφουμε τη δομή και τη χρήση των εν λόγω οντολογιών.

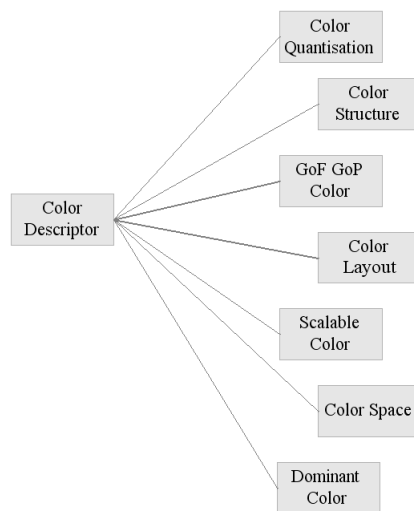
3.3.1 Οντολογία Οπτικών Περιγραφών

Η οντολογία οπτικών περιγραφών (Visual Descriptor Ontology - VDO) μοντελοποιεί έννοιες και ιδιότητες που περιγράφουν οπτικά χαρακτηριστικά ενός αντικειμένου. Παρόλο που η κατασκευή της VDO είναι στενά συνδεδεμένη με τις προδιαγραφές που καθορίζονται στο πρότυπο MPEG-7, χρειάστηκε να γίνουν αρκετές τροποποιήσεις για να μετατραπεί το XML σχήμα του προτύπου στην οντολογία οπτικών περιγραφών VDO.

Η έννοια VDO:VISUALDESCRIPTOR βρίσκεται στην κορυφή της οντολογίας οπτικών περιγραφών και σε αυτήν υπάγονται όλες οι έννοιες των μοντέλων των οπτικών περιγραφών. Συνίσταται βασικά από έξι υπό-έννοιες, μία για κάθε κατηγορία που ορίζει το πρότυπο MPEG-7, και συγκεκριμένα το χρώμα, το σχήμα, η υφή, η κίνηση, ο εντοπισμός και οι βασικοί περιγραφείς. Ως παράδειγμα, στο σχήμα 3.2 απεικονίζονται οι υπό-έννοιες που υπάγονται από την έννοια COLORDESCRIPTOR. Κάθε μία από τις υπαγόμενες αυτές έννοιες περιλαμβάνει έναν αριθμό από σχετικούς περιγραφείς οι οποίοι με τη σειρά τους ορίζονται ως έννοιες στην VDO.

3.3.2 Οντολογία Δομής Πολυμέσων

Η οντολογία δομής πολυμέσων (Multimedia Structure Ontology - MSO) μοντελοποιεί τις κύριες πολυμεσικές έννοιες του σχήματος περιγραφής πολυμέσων του προτύπου



Σχήμα 3.2: Ιεραρχία οντολογίας οπτικών περιγραφών.

MPEG-7 [106] και τις αμοιβαίες σχέσεις, όπως είναι για παράδειγμα η αποδόμηση (decomposition). Εντός του MPEG-7, το πολυμεσικό περιεχόμενο χωρίζεται σε πέντε είδη: Εικόνα, βίντεο, ήχος, οπτικοακουστικό και πολυμέσα. Το πρότυπο προβλέπει για κάθε ένα από τα πέντε αυτά είδη τις δικές του κλάσεις τμημάτων προκειμένου να περιγράψει τη δομή του καθενός στο χωροχρόνο. Ένα τμήμα ορίζεται ως μία κλάση που περιγράφει τα χωρικά ή/και χρονικά όρια ενός πολυμεσικού περιεχομένου. Πιο συγκεκριμένα για το βίντεο, προβλέπονται οι κλάσεις “τμήμα βίντεο”, “κινούμενη περιοχή”, “στατική περιοχή”, “μωσαϊκό” οι οποίες μπορούν να προκύψουν από χωρική, χρονική ή χωροχρονική κατάτμηση του βίντεο.

3.3.3 Οντολογία Πυρήνα και Χωρο-χρονικές Επεκτάσεις

Ο ρόλος της οντολογίας πυρήνα (core ontology) στο συνολικό πλαίσιο μίας οντολογικής βάσης γνώσης, είναι να λειτουργεί ως σημείο αναφοράς για τις επιμέρους οντολογίες. Πιο συγκεκριμένα, για αυτόν το σκοπό, επιλέξαμε την οντολογία DOLCE [42]. Η DOLCE είναι μία ευρέως χρησιμοποιημένη και τεκμηριωμένη οντολογία πυρήνα, με αυστηρά ορισμένα αξιώματα, διατηρώντας ταυτόχρονα μία δομή επαναχρησιμοποιούμενων και ευρέως χρησιμοποιούμενων κατηγοριών και εννοιών.

Αν και η DOLCE παρέχει τα μέσα για την αναπαράσταση χωρο-χρονικών χαρακτηριστικών, δεν επιτρέπει την αναπαράσταση σύνθετων τοπολογικών και κατευθυντικών σχέσεων μεταξύ χρονικών ή χωρικών περιοχών. Οι κατευθυντικές χωρικές σχέσεις είναι απαραίτητες για να περιγράψουν πως περιοχές της εικόνας είναι τοποθετημένες στο 2D χώρο (την απόλυτη αλλά και τη σχετική τους θέση, όπως π.χ. ΨΗΛΑ ή ΔΕΞΙΑΑΠΟ). Οι τοπολογικές χωρικές σχέσεις περιγράφουν πως συνδέονται τα περιγράμματα των περιοχών, όπως για παράδειγμα ΑΚΟΥΜΠΑΕΙ ή ΕΠΙΚΑΛΥΠΤΕΙ. Με παρόμοιο τρόπο, οι χρονικές σχέσεις [3] χρησιμοποιούνται για να αναπαραστήσουν τις σχέσεις των περιοχών στη διάσταση του χρόνου. Επιπρόσθετα, επεκτείναμε την έννοια ΠΕΡΙΟΧΗ της οντολογίας DOLCE, ώστε να υποστηρίξουμε τοπολογικές και κατευθυντικές σχέσεις μεταξύ περιοχών διαφορετικού τύπου, όπως για παράδειγμα των τύπων “Χρονική Περιοχή” και “2D Περιοχή”, οι οποίες είναι χρήσιμες καθώς

αποτελούν συχνά το αποτέλεσμα διαδικασιών ανάλυσης πολυμέσων.

3.3.4 Οντολογία Θεματικού Πεδίου

Ο ρόλος των οντολογιών θεματικού πεδίου (domain ontologies) είναι η μοντελοποίηση του πολυμεσικού υλικού σε σχέση με συγκεκριμένα πεδία/σενάρια πραγματικού κόσμου, όπως για παράδειγμα “καλοκαιρινές διακοπές”, “τένις”, κ.ά. Όλες οι οντολογίες θεματικού πεδίου βασίζονται απευθείας ή είναι ευθυγραμμισμένες με την οντολογία πυρήνα DOLCE, έτσι ώστε να συνδέονται με έννοιες υψηλού επιπέδου, εξασφαλίζοντας διαλειτουργικότητα (interoperability) μεταξύ τους.

Οι οντολογίες θεματικού πεδίου ορίζονται με τέτοιο τρόπο ώστε να παρέχουν ένα γενικό μοντέλο του πεδίου που περιγράφουν, λαμβάνοντας όμως υπόψη και τη συγκεκριμένη οπτική ενός χρήστη ή απαίτηση μίας εφαρμογής. Η μοντελοποίηση ενός θεματικού πεδίου από τη μία εξυπηρετεί τη λογική και σημασιολογική περιγραφή του από τις έννοιες και τις σχέσεις αυτών που το αποτελούν, ενώ από την άλλη θα πρέπει να είναι εφικτή η ανάκτηση τέτοιων εννοιών από αυτόματα εργαλεία ανάλυσης πολυμέσων. Με άλλα λόγια, οι έννοιες που περιγράφονται σε μία οντολογία θεματικού πεδίου πρέπει να μπορούν να εξαχθούν αυτόματα, ενώ θα πρέπει να παραμένουν κατανοητές από έναν άνθρωπο (σε αντιδιαστολή με αριθμητικές αναπαραστάσεις όπως τα διανύσματα χαρακτηριστικών).

3.3.5 Προσέγγιση με Πρότυπα Στιγμιότυπα

Η περιγραφή των χαρακτηριστικών των εννοιών με σκοπό τη χρήση τους στην ανάλυση, φυσιολογικά οδηγεί σε ένα δίλημμα μοντελοποίησης των χαρακτηριστικών αυτών. Το δίλημμα προκύπτει καθώς η χρήση των τιμών χαρακτηριστικών ως έννοιες ή ιδιότητες εννοιών δεν είναι εφικτή απευθείας χωρίς να αποφύγουμε δευτέρου βαθμού μοντελοποίηση. Στη δική μας προσέγγιση, η βάση γνώσης εμπλουτίζεται με στιγμιότυπα εννοιών, που ανήκουν στις οντολογίες θεματικών πεδίων, τα οποία εξυπηρετούν ως πρότυπα για αυτές τις έννοιες. Αυτό το επιτυγχάνουμε με την εισαγωγή επιπλέον εννοιών (του τύπου VDOEXT:PROTOTYPE) που συγκεκριμενοποιούν (instantiate) τις αφηρημένες έννοιες συνδέοντάς τις με έννοιες τις οντολογίας οπτικών περιγραφών. Οι έννοιες τις οντολογίας περιγραφών περιέχουν τις τιμές των χαρακτηριστικών των εκάστοτε πρότυπων στιγμιότυπων, επιτρέποντας μία ρεαλιστική και εύκολα επεκτάσιμη προσέγγιση, διατηρώντας παράλληλα καθαρή την οντολογία θεματικών πεδίων από τα πρότυπα στιγμιότυπα.

Η προτεινόμενη υποδομή οντολογίας χρησιμοποιεί ως γλώσσα αναπαράστασης τη γλώσσα Resource Description Framework Schema (RDFs). Η RDFs παρέχει ικανοποιητικό αριθμό βασικών τύπων (primitives) για να μπορέσει να καθορίσει τα μοντέλα των εννοιών των οντολογιών θεματικών πεδίων. Αντίθετα, άλλα μέρη της συνολικής οντολογίας, όπως είναι για παράδειγμα η οντολογία πυρήνα DOLCE, είναι κωδικοποιημένα στη γλώσσα OWL ή σε κάποια διάλεκτο αυτής. Η απόφαση μας να αναπαραστήσουμε τις οντολογίες θεματικών πεδίων με RDFS οφείλεται στο γεγονός ότι η χρήση μίας τόσο εκφραστικής γλώσσας όπως η OWL θα απαιτούσε πολύ περισσότερο πολύπλοκους μηχανισμούς ανάλυσης και διαδικασίες συμπερασματολογίας. Κάτι τέτοιο, και δεδομένου του όγκου των στιγμιότυπων που θα απαιτηθεί να επεξεργαστούν ανά πάσα στιγμή, θα αύξανε υπερβολικά την πολυπλοκότητα και θα καθιστούσε ανέφικτη την περαιτέρω εξέλιξη της ανάλυσης μίας εικόνας ή μίας ακολουθίας βίντεο.

3.4 Εξαγωγή και Αποθήκευση Χαρακτηριστικών

Στη συγκεκριμένη ενότητα θα περιγράψουμε τη μεθοδολογία εξαγωγής χαρακτηριστικών περιοχών της εικόνας και το ταίριασμα αυτών με πρότυπα στιγμιότυπα εννοιών με τελικό σκοπό την ταξινόμησή τους.

Καταρχήν, η εικόνα κατατέμνεται σε έναν αριθμό περιοχών με τη βοήθεια ενός αλγορίθμου κατάτμησης, ούτως ώστε να καταστεί δυνατή η ανάλυση με βάση τις περιοχές και όχι τις τιμές των εικονοστοιχείων. Στη συνέχεια εξάγουμε ορισμένους τυπικούς οπτικούς περιγραφείς του προτύπου MPEG-7 και τους αποθηκεύουμε σε ένα σχεσιακό γράφο χαρακτηριστικών *ARG* όπως αυτός περιγράφηκε στην παράγραφο 2.3.1. Στη συνέχεια ακολουθεί μία διαδικασία ανάκτησης των τιμών των χαρακτηριστικών πρότυπων στιγμιότυπων από τη βάση γνώσης με σκοπό το ταίριασμα και την ταξινόμηση των περιοχών. Τα πρότυπα στιγμιότυπα αποτελούν μέρος της συνολικής βάσης γνώσης που περιγράφηκε στην ενότητα 3.3, η οποία αρχικά χρησιμεύει για την αποθήκευση εννοιών και χαρακτηριστικών συνδεδεμένων με την εικόνα και τις περιοχές τις, αλλά μπορεί να χρησιμεύσει επίσης και για την εξαγωγή συμπερασμάτων με σκοπό την περιγραφή της σε ένα ανώτερα σημασιολογικά επίπεδο.

3.4.1 Αρχική Κατάτμηση

Πολλές τεχνικές ανάλυσης εικόνων εφαρμόζονται σε ομάδες εικονοστοιχείων ή αλλιώς περιοχές της εικόνας, οι οποίες μοιράζονται κάποιο ή κάποια κοινά χαρακτηριστικά. Έτσι είναι απαραίτητο στο στάδιο προ-επεξεργασίας να εφαρμοστεί κάποιος αλγόριθμος κατάτμησης (περισσότερα στο κεφάλαιο 4) για να δημιουργήσει ομογενείς (ως προς κάποιο κριτήριο) περιοχές. Αυτή την προσέγγιση επιλέξαμε και εμείς και χρησιμοποιήσαμε μία παραλλαγή [1] του διαδομένου αλγορίθμου αναδρομικά ελάχιστου συνδετικού δένδρου (recursive shortest spanning tree - RSST) [82]. Αυτό το στάδιο προ-επεξεργασίας παράγει ένα σύνολο συνδεδεμένων περιοχών, έτσι ώστε η περαιτέρω ανάλυση να γίνει στις περιοχές αυτές και όχι στο επίπεδο των εικονοστοιχείων. Δύο παραδείγματα κατάτμησης απεικονίζονται στο σχήμα 3.4, όπου η αρχική εικόνα έχει κατατμηθεί σε μερικές περιοχές, κάθε μία εκ των οποίων απεικονίζεται με μία σταθερή γκρι τιμή (κανονικοποιημένη για να είναι πιο ευδιάκριτη η εικόνα).

3.4.2 Εξαγωγή και Ταίριασμα Χαρακτηριστικών

Προκειμένου να αναπαραστήσουμε αποτελεσματικά την οπτική πληροφορία που περιλαμβάνεται σε μία περιοχή της εικόνας, επιλέξαμε ορισμένους οπτικούς περιγραφείς που προτείνει το πρότυπο MPEG-7. Πιο συγκεκριμένα επιλέξαμε έναν περιγραφέα για τρία βασικά οπτικά χαρακτηριστικά, το χρώμα, την υφή και το σχήμα. Οι οπτικοί περιγραφείς που χρησιμοποιήσαμε για την αναπαράσταση των χαμηλού-επιπέδου χαρακτηριστικών της κάθε περιοχής είναι ο περιγραφέας κύριων χρωμάτων (*DC*), ο περιγραφέας ομοιογενούς υφής (*HT*) [112] και ο περιγραφέας σχήματος με βάση το σχήμα (*RS*) [18]. Για τη μοντελοποίηση και αποθήκευση των υψηλού-επιπέδου χαρακτηριστικών χρησιμοποιήσαμε ασαφή σύνολα πιθανών συνδεδεμένων με τις περιοχές εννοιών.

3.4.2.1 Περιγραφέας Κύριων Χρωμάτων

Ο περιγραφέας κύριων χρωμάτων (DC) [69] αποτελεί μία πολύ συμπαγή και αποτελεσματική περιγραφή της χρωματικής κατανομής μίας περιοχής. Αποτελείται από τα κύρια χρώματα (μέση τιμή όπως αναπαριστώνται στον τρισδιάστατο χώρο χρώματος), τα ποσοστά κάλυψης ως προς το εμβαδόν της περιοχής, τη διακύμανση τους και τη χωρική τους συνοχή. Ο DC είναι ιδιαίτερα χρήσιμος για εφαρμογές ανάκτησης εικόνων με βάση το χρώμα καθώς και ταξινόμηση περιοχών.

Για τον υπολογισμό του περιγραφέα κύριων χρωμάτων, τα χρώματα που είναι παρόντα σε μια δεδομένη εικόνα ή μια περιοχή ενδιαφέροντος πρώτα συσταδοποιούνται και στη συνέχεια υπολογίζονται οι μέσες τιμές της κάθε συστάδας και προαιρετικά υπολογίζονται επίσης οι διακυμάνσεις των κύριων χρωμάτων. Επίσης υπολογίζεται και το ποσοστό κάλυψης της περιοχής από την εν λόγω συστάδα. Τέλος υπολογίζεται μια τιμή χωρικής συνοχής που βοηθά στη διαφοροποίηση μεταξύ των μεγάλων ενιαίων περιοχών χρώματος εναντίον των χρωμάτων που είναι εξαπλωμένα σε όλη την εικόνα. Ο αριθμός των κυρίων χρωμάτων μπορεί να ποικίλει από περιοχή σε περιοχή και ένα μέγιστο οκτώ κύριων χρωμάτων μπορεί να χρησιμοποιηθεί για να περιγράψει την εικόνα/περιοχή.

$$DC = [\{c_i, v_i, p_i\}, s], i = 1..N \quad (3.1)$$

όπου c_i είναι το i -οστό κύριο χρώμα, v_i η διακύμανση του χρώματος, p_i το ποσοστό κάλυψης του i -οστού κύριου χρώματος, s η χωρική συνοχή και N ο αριθμός κυρίων χρωμάτων ο οποίος είναι οκτώ. Κάθε περιοχή μπορεί να έχει το πολύ οκτώ κύρια χρώματα, παρόλα αυτά, πειράματα έχουν δείξει ότι τρία με τέσσερα κύρια χρώματα είναι αρκετά για να περιγράψουν ικανοποιητικά τη χρωματική κατανομή μίας περιοχής.

Η μετρική απόστασης $d_{DC}(DC_a, DC_b)$ ¹ για δύο περιγραφείς DC_a και DC_b δύο περιοχών a, b ορίζεται από την ακόλουθη σχέση:

$$d_{DC}(a, b) = \sqrt{\sum_{i=1}^{N_1} p_{ai}^2 + \sum_{j=1}^{N_2} p_{bj}^2 - \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} 2\alpha_{ai,bj} p_{ai} p_{bj}} \quad (3.2)$$

Ο συντελεστής $\alpha_{ai,bj}$ είναι ο συντελεστής ομοιότητας δύο χρωμάτων c_1, c_2 και ορίζεται ως εξής:

$$\alpha_{1,2} = \begin{cases} 1 - \frac{\|c_1 - c_2\|}{\|c_1 - c_2\|_{max}}, & \text{αν } \|c_1 - c_2\| \leq T_d \\ 0, & \text{αν } \|c_1 - c_2\| > T_d \end{cases} \quad (3.3)$$

3.4.2.2 Περιγραφέας Ομοιογενούς Υφής

Ο περιγραφέας ομοιογενούς υφής HT [69] παρέχει έναν ποσοτικό χαρακτηρισμό των περιοχών ομοιογενούς υφής για ανάκτηση ομοιότητας με βάση την υφή. Βασίζεται στον υπολογισμό των τοπικών χωρικών και συχνοτικών στατιστικών της υφής. Η περιοχή, αρχικά, φιλτράρεται με μια σειρά ευαίσθητων φίλτρων προσανατολισμού και κλίμακας και υπολογίζεται η μέση και σταθερή απόκλιση των φιλτραρισμένων αποτελεσμάτων στο χώρο της συχνότητας. Η υπολογιστική πολυπλοκότητα αυτού του περιγραφέα μπορεί να μειωθεί σημαντικά με τον υπολογισμό των τιμών στην περιοχή συχνότητας παρά στη χωρική περιοχή.

¹Για λόγους απλότητας και αναγνωσιμότητας θεωρούμε για την απόσταση δύο περιγραφέων ισοδύναμους τους συμβολισμούς $d_{DC}(DC_a, DC_b) \equiv d_{DC}(a, b)$ (παρομοίως και για τις d_{HT}, d_{RS}).

Πιο συγκεκριμένα ο περιγραφέας ομοιογενούς υφής ορίζεται ως:

$$HT \equiv [avg, std, e_1, \dots, e_{30}, d_1, \dots, d_{30}] \quad (3.4)$$

Όπου avg είναι η μέση τιμή και std η τυπική απόκλιση των τιμών της φωτεινότητας της περιοχής, αντίστοιχα. e_i και d_i είναι η μέση ενέργεια και η τυπική απόκλιση της ενέργειας της απόκρισης για κάθε ένα από τα τριάντα φίλτρα ($i \in [1, \dots, 30]$).

Τέλος ορίζεται η απόσταση δύο περιγραφέων HT_a και HT_b δύο περιοχών a, b με την ακόλουθη σχέση:

$$d_{HT}(a, b) = \sum_{i=1}^{N_{HT}=62} \left| \frac{HT_a(i) - HT_b(i)}{\sigma_i} \right| \quad (3.5)$$

όπου σ_i είναι μία παράμετρος κανονικοποίησης για κάθε συχνοτικό κανάλι.

3.4.2.3 Περιγραφέας Σχήματος με βάση την Περιοχή

Ο περιγραφέας σχήματος με βάση την περιοχή [18] εκφράζει τη κατανομή εικονοστοιχείων μέσα σε μια 2Δ περιοχή αντικειμένου, μπορεί να περιγράψει σύνθετα αντικείμενα που αποτελούνται από πολλαπλάσιες ασύνδετες περιοχές καθώς επίσης και απλά αντικείμενα με ή χωρίς τρύπες. Συνυπολογίζει όλα τα εικονοστοιχεία που αποτελούν το σχήμα, δηλαδή το περίγραμμα και τα εσωτερικά εικονοστοιχεία. Δεν επηρεάζεται από περιστροφή και κλιμάκωση. Δύο σημαντικά χαρακτηριστικά γνωρίσματα αυτού του περιγραφέα είναι τα α) ότι δίνει ένα συμπαγή και αποδοτικό τρόπο περιγραφής των ιδιοτήτων πολλαπλών ασύνδετων περιοχών ταυτόχρονα και β) ότι είναι εύρωστος στο θόρυβο κατάτμησης, π.χ., στον κρουστικό θόρυβο (salt and pepper).

Πιο συγκεκριμένα, εξάγονται οι συντελεστές του γωνιακού ακτινικού μετασχηματισμού [18] (angular radial transformation - ART):

$$RS \equiv [M_i] \quad (3.6)$$

όπου M_i είναι οι κανονικοποιημένοι και κβαντισμένοι ART συντελεστές.

Το πρότυπο MPEG-7 δεν καθορίζει μία συνάρτηση απόστασης μεταξύ δύο περιγραφέων RS , οπότε υιοθετήσαμε την κλασική Ευκλείδεια απόσταση:

$$d_{RS}(a, b) = \sqrt{\sum_{i=1}^{|RS|} (RS_a(i) - RS_b(i))^2} \quad (3.7)$$

3.4.3 Ανάκτηση Πρότυπων Στιγμιότυπων

Για την ανάκτηση των πρότυπων στιγμιότυπων και των περιγραφέων των, χρησιμοποιούμε τη μηχανή OntoBroker η οποία μπορεί να επικοινωνεί με τη βάση γνώσης μέσω κατάλληλων ερωτημάτων, ενώ επιπρόσθετα υποστηρίζει οντολογίες αποθηκευμένες σε RDFs. Για τους σκοπούς της ανάλυσης, το OntoBroker καλείται να φορτώσει τις οντολογίες θεματικής περιοχής, όπου είναι δηλωμένες έννοιες υψηλού επιπέδου, την οντολογία οπτικών περιγραφέων η οποία περιλαμβάνει τους ορισμούς των οπτικών περιγραφέων χαμηλού επιπέδου, καθώς και τα πρότυπα στιγμιότυπα τα οποία συνδέουν τις έννοιες με τους περιγραφείς, δηλαδή όλη τη βάση γνώσης του συστήματός μας. Όπως αναφέρθηκε και προηγουμένως, η επικοινωνία του συστήματος ανάλυσης με τη

βάση γνώσης επιτυγχάνεται μέσω κατάλληλων ερωτημάτων, με τα οποία επιτυγχάνεται η ανάκτηση τιμών συγκεκριμένων οπτικών περιγραφών για οποιαδήποτε έννοια. Το OntoBroker χρησιμοποιεί τη γλώσσα λογικής πλαισίου (f-logic). Η γλώσσα f-logic είναι μία γλώσσα αναπαράστασης η οποία μπορεί να χρησιμοποιηθεί για να μοντελοποιηθούν οντολογίες, καθώς επίσης και γλώσσα ερωτημάτων (query language), η οποία χρησιμοποιείται για την επικοινωνία με τη βάση γνώσης του Ontobroker.

Το Ontobroker προσφέρει επικοινωνία (API) με ορισμένες γλώσσες προγραμματισμού, όπως η JAVA και η C++. Καθώς η δεύτερη προτιμήθηκε για την υλοποίηση των αλγορίθμων ανάλυσης, χρησιμοποιήθηκε και το αντίστοιχο API, έτσι ώστε να εξασφαλιστεί η σύνδεση του συστήματος ανάλυσης μέσω web service με τον εξυπηρετητή στον οποίο τρέχει το Ontobroker και είναι φορτωμένη η βάση γνώσης. Προκειμένου να ανακτήσουμε τα πρότυπα στιγμιότυπα από τη βάση γνώσης με επιτυχία, υπάρχουν μερικοί εναλλακτικοί τρόποι, ωστόσο σε κάθε περίπτωση χρειάζεται να θέσουμε μία σειρά από ερωτήματα. Ακολουθούν ορισμένα παραδείγματα και η βήμα προς βήμα επεξήγησή αυτών για να καταστεί σαφές πως μπορεί να επιτευχθεί το παραπάνω. Ας υποθέσουμε ότι θέλουμε να ανακτήσουμε όλους τους διαθέσιμους οπτικούς περιγραφείς της έννοιας “μπάλα τένις”. Καταρχήν, με το παρακάτω απλό ερώτημα σε f-Logic στο Ontobroker αναχτούμε όλα τα πρότυπα παραδείγματα της συγκεκριμένης έννοιας:

```
FORALL X<- "http://www.acemedia.org/ontologies/tennis"#Ball.
```

Το αποτέλεσμα του παραπάνω ερωτήματος είναι:

```
X = "http://www.acemedia.org/ontologies/VDO-EXT"#Ball_Prototype_1"
```

Συνεχίζοντας, το επόμενο βήμα είναι να ζητήσουμε όλους τους διαθέσιμους οπτικούς περιγραφείς (όπως αυτοί ορίζονται στην VDO) για το πρότυπο παράδειγμα Ball_Prototype_1:

```
FORALL X<- "http://www.acemedia.org/ontologies/VDO-EXT"#Ball_Prototype_1  
["http://www.acemedia.org/ontologies/VDO-EXT"#hasDescriptor->>>X].
```

Το αποτέλεσμα του τελευταίου ερωτήματος είναι ότι το Ball_Prototype_1 είναι συσχετισμένο με τέσσερις οπτικούς περιγραφείς:

```
X="http://www.acemedia.org/ontologies/VDO"#VDE_INST_110147216871832572097"  
X="http://www.acemedia.org/ontologies/VDO"#VDE_INST_110147216053132572097"  
X="http://www.acemedia.org/ontologies/VDO"#VDE_INST_110147217659332572097"  
X="http://www.acemedia.org/ontologies/VDO"#VDE_INST_110147218634332572097"
```

Έπειτα μπορούμε να ανακτήσουμε τις τιμές του κάθε ένα οπτικού περιγραφέα, εφόσον αναγνωρίσουμε τον τύπο του (πχ Dominant Color, Edge Histogram, κλπ.):

```
FORALL
```

```
X<- "http://www.acemedia.org/ontologies/VDO"#VDE_INST_110147216871832572097:X.
```

Το παραπάνω ερώτημα επιστρέφει όλες τις έννοιες της οντολογίας οπτικών περιγραφών στις οποίες ανήκει το συγκεκριμένο παράδειγμα:

```
X = "http://www.acemedia.org/ontologies/VDO"#DEFAULT_ROOT_CONCEPT"  
X_□=□ "http://www.acemedia.org/ontologies/VDO"#ColorDescriptor"  
X = "http://www.acemedia.org/ontologies/VDO"#VisualDescriptor"  
X_□=□ "http://www.acemedia.org/ontologies/VDO"#ScalableColorDescriptor"
```

Εφόσον βρήκαμε ότι το παράδειγμα περιγραφέα VDE_INST_110147216871832572097 αναφέρεται στον περιγραφέα ScalableColorDescriptor, μπορούμε να ανακτήσουμε τις πραγματικές του τιμές:

FORALL

X,Y←http://www.acemedia.org/ontologies/VDO#VDE_INST_110147216871832572097[X→Y].

Οι τιμές που επιστρέφονται είναι:

X = "http://www.acemedia.org/ontologies/VDO"#numberOfCoefficients",

Y_□=□0

X_□=□"http://www.acemedia.org/ontologies/VDO"#numberOfBitPlanesDiscarded"

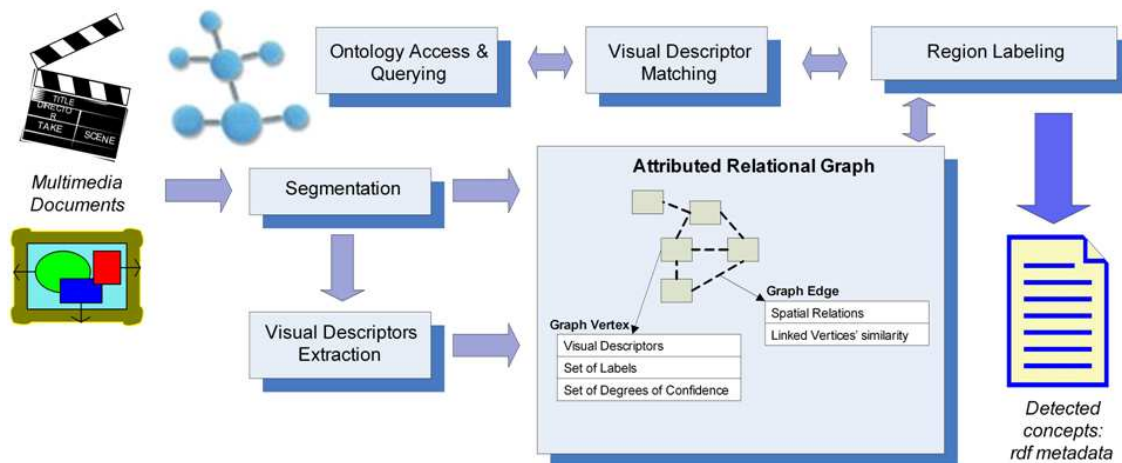
Y = 6

X = "http://www.acemedia.org/ontologies/VDO"#coefficients",

Y_□=□"0 0 0 1 1 1 1 1 1 1 1 1 1 1"

Η παραπάνω διαδικασία χρειάζεται να επαναληφθεί για όλους τους οπτικούς περιγραφείς που είναι συνδεδεμένοι με το παράδειγμα Ball_Prototype_1 και συνεπώς για όλες τις έννοιες της οντολογίας θεματικής περιοχής (η θεματική περιοχή Τένις στο παραπάνω παράδειγμα) για τις οποίες ενδιαφέρεται να εντοπίσει το σύστημα ανάλυσης. Όλες οι τιμές αποθηκεύονται σε κατάλληλες εσωτερικές δομές της C++ και συνιστούν την a-priori γνώση την οποία διαχειρίζεται το σύστημα ανάλυσης κατά τη διάρκεια του ταιριάσματος περιγραφών προκειμένου να χαρακτηρίσει σημασιολογικά το περιεχόμενο.

3.4.4 Αρχικοποίηση Σχεσιακού Γράφου Χαρακτηριστικών ARG



Σχήμα 3.3: Αρχιτεκτονική ανάλυσης πολυμέσων με βοήθεια γνώσης.

Για την αξιοποίηση της προαναφερθείσας αναπαράστασης γνώσης και τη δημιουργία σημασιολογικών μεταδεδομένων που να βασίζονται στο περιεχόμενο, έχει δημιουργηθεί μια πλατφόρμα για τον έλεγχο και την αξιολόγηση τόσο των οντολογιών που δημιουργήθηκαν, όσο και των μεταδεδομένων που εξάγονται. Η πλατφόρμα αυτή ονομάζεται ΚΑΑ (knowledge assisted analysis) και η αρχιτεκτονική της φαίνεται στο σχήμα 3.3.

Ο γράφος ARG αρχικοποιείται πάνω σε μία υπερκατάτμηση της εικόνας βασισμένη σε αλγόριθμο κατάτμησης με βάση το χρώμα και το σχήμα [1]. Για κάθε μία περιοχή που δημιουργείται εξάγουμε τους περιγραφείς DC, HT και RS (εξισώσεις (3.1), (3.4), (3.6)) και συνθέτουμε ένα διάνυσμα χαρακτηριστικών $D_a = [DC_a HT_a RS_a]$ για κάθε περιοχή a και αποθηκεύεται στον αντίστοιχο κόμβο v_a του ARG. Στον κάθε κόμβο αποθηκεύονται οι οπτικοί περιγραφείς της αντίστοιχης περιοχής, καθώς και ορισμένες

άλλες πληροφορίες (δυναμική μάσκα, περίγραμμα, κλπ). Η ύπαρξη ενός συνδέσμου ανάμεσα σε δύο κόμβους του γράφου σημαίνει ότι οι περιοχές που αντιστοιχούν σε αυτούς τους κόμβους είναι χωρικά γειτονικές. Παράλληλα, αποθηκεύεται η απόσταση μεταξύ των δύο περιοχών με βάση τους οπτικούς περιγραφείς της, καθώς και τα κοινά στοιχεία του περιγράμματος των δύο περιοχών.

Ο βαθμός ομοιότητας (εναλλακτικά, η απόσταση) μεταξύ των περιοχών μπορεί να υπολογιστεί με δύο τρόπους. Ο πρώτος είναι με τον υπολογισμό των αποστάσεων με τη χρήση των μετρικών συναρτήσεων σύμφωνα του προτύπου MPEG-7 (εξισώσεις (3.2), (3.5), (3.7)). Στη συνέχεια η μοναδική απόσταση μεταξύ των δύο περιοχών υπολογίζεται από ένα άθροισμα με βάρη, τα οποία προκύπτουν πειραματικά, έπειτα από την αξιολόγηση των αποτελεσμάτων με βάση ένα σύνολο εικόνων εκπαίδευσης. Ο δεύτερος τρόπος είναι με τη χρήση ενός νευρωνικού δικτύου back-propagation. Το δίκτυο δέχεται σαν είσοδο τις περιγραφές των περιοχών και έχει ως έξοδο την κανονικοποιημένη απόστασή τους. Επειδή η γνώση ότι δύο περιοχές είναι γειτονικές δεν είναι αρκετή για την αποτελεσματική ανάλυση, εξάγονται συμπληρωματικά οι χωρικές σχέσεις μεταξύ των περιοχών. Οι σχέσεις αυτές μπορεί να είναι σχετικές, όπως για παράδειγμα “η περιοχή 1 βρίσκεται πάνω από την περιοχή 2”, ή απόλυτες, όπως “η περιοχή 1 βρίσκεται πάνω από όλες τις περιοχές”.

Με τον ίδιο τρόπο όπως γίνεται ο υπολογισμός της απόστασης μεταξύ δύο περιοχών, γίνεται και η αυτός μεταξύ μιας περιοχής και μιας έννοιας. Όλες οι πιθανές έννοιες για μια περιοχή αποθηκεύονται στον αντίστοιχο κόμβο του γράφου και κάθε μια συνοδεύεται από ένα βαθμό ομοιότητας ανάμεσα στην έννοια και την περιοχή. Οι έννοιες που έχουν βαθμό ομοιότητας κάτω από ένα προκαθορισμένο κατώφλι απορρίπτονται για την περαιτέρω ανάλυση. Στην περίπτωση που δύο ή περισσότερες γειτονικές περιοχές έχουν προσδιοριστεί ότι ανήκουν στην ίδια έννοια, θεωρούμε ότι αποτελούν μέρος μιας μεγαλύτερης περιοχής που χωρίστηκε από λάθος της κατάτμησης. Οι περιοχές ενώνονται και παράλληλα ανανεώνεται ο γράφος καθώς και όλες οι πληροφορίες που έχουν αποθηκευθεί σε αυτόν. Έτσι, για παράδειγμα, εξάγονται οι οπτικοί περιγραφείς για τις νέες περιοχές που προέκυψαν, υπολογίζονται εκ νέου οι αποστάσεις μεταξύ περιοχών και εννοιών κλπ.

Κάθε φορά που μια εικόνα δίνεται στην πλατφόρμα για ανάλυση, η βάση γνώσης που έχει κατασκευαστεί εκ των προτέρων χρησιμοποιείται για την αντιστοίχιση περιοχών της εικόνας με έννοιες. Αυτό επιτυγχάνεται με τη σύγκριση των MPEG-7 περιγραφών των περιοχών που προκύπτουν μετά την κατάτμηση με τις αντίστοιχες των πρότυπων στιγμιότυπων των εννοιών που έχουν κωδικοποιηθεί στις οντολογίες θεματικών πεδίων. Για την ανάκτηση των στιγμιότυπων χρησιμοποιείται η μηχανή Ontobroker². Η πλατφόρμα Ontobroker επιτρέπει το φόρτωμα των οντολογιών που είναι κωδικοποιημένες στη γλώσσα αναπαράστασης RDFs και πιο συγκεκριμένα της VDO οντολογίας που περιλαμβάνει τους ορισμούς των οπτικών περιγραφέων που χρησιμοποιήσαμε, καθώς και της οντολογίας που περιλαμβάνει κωδικοποιημένη γνώση της σημασιολογικής περιοχής στην οποία ανήκει η προς ανάλυση εικόνα. Στην περίπτωση που για κάποια έννοια χρησιμοποιείται ο ορισμός της σε κάποια προγενέστερη οντολογία, όπως για παράδειγμα η DOLCE οντολογία για την έννοια ΑΝΘΡΩΠΟΣ, το Ontobroker μπορεί να τη φορτώσει παράλληλα με τις προηγούμενες.

²<http://www.ontoprise.de/en/home/products/ontobroker>

3.5 Κατασκευή Σημασιολογικών Μεταδεδομένων

Ο σκοπός της ανάλυσης είναι να εξάγει χαρακτηριστικά υψηλού επιπέδου, τα οποία να γίνονται αντιληπτά από το χρήστη. Παράλληλα με αυτό είναι αναγκαίο να δημιουργεί σημασιολογικά μεταδεδομένα τα οποία περιγράφουν με τυποποιημένο τρόπο το πολυμεσικό περιεχόμενο. Στην αναγκαιότητα των σημασιολογικών μεταδεδομένων αναφερθήκαμε στην εισαγωγή της διατριβής, καθώς παρέχουν τη δυνατότητα για εύκολη αναζήτηση, πλοήγηση και ανάκτηση του πολυμεσικού υλικού.

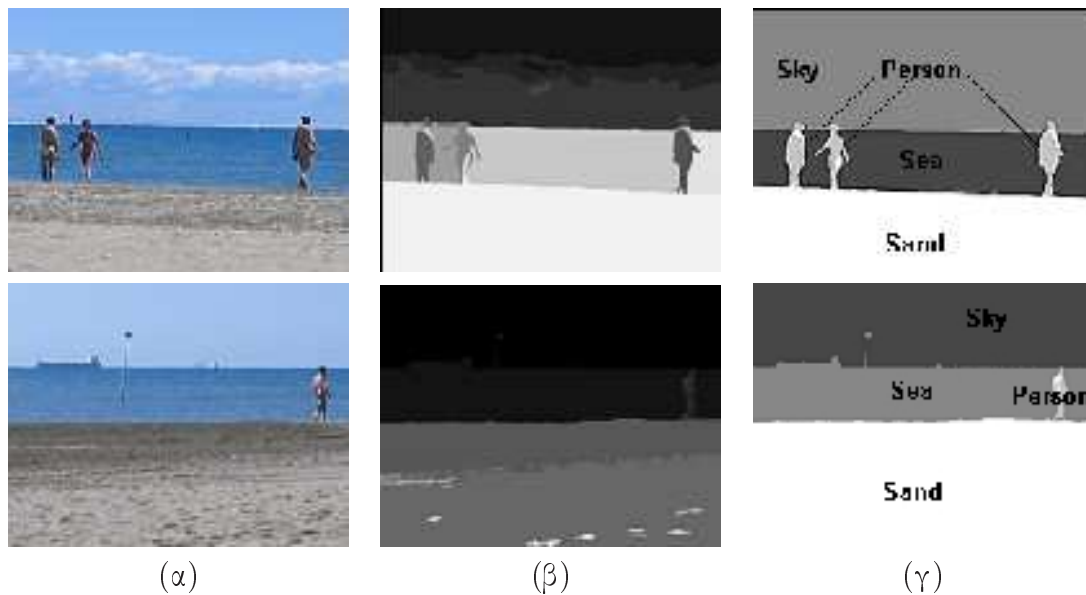
Για την αναπαράσταση των σημασιολογικών μεταδεδομένων επιλέχθηκε η γλώσσα περιγραφής RDFs, καθώς θεωρήθηκε ότι η εκφραστικότητά της είναι επαρκής και επιπλέον διευκολύνεται η διασύνδεση των μεταδεδομένων με την οντολογία MSO που περιγράφει τη δομή του πολυμεσικού υλικού. Η δομή αυτή είναι κατά κύριο λόγο βασισμένη στην MSO, η οποία μοντελοποιεί τις κύριες πολυμεσικές έννοιες του σχήματος περιγραφής πολυμέσων του προτύπου MPEG-7 [106] και τις αμοιβαίες σχέσεις, όπως είναι για παράδειγμα η αποδόμηση.

Πιο συγκεκριμένα, η ιεραρχική αναπαράσταση των σημασιολογικών μεταδεδομένων συνίσταται από τα ακόλουθα τέσσερα επίπεδα: Τμήματος, ιδιότητες τμημάτων, οπτικής περιγραφής και σημασιολογικών μεταδεδομένων. Το πρώτο επίπεδο αναπαριστά το είδος της περιγραφόμενης οντότητας και παρέχει την απαραίτητη χωροχρονική αποδόμηση, για παράδειγμα αν πρόκειται για μία περιοχή της εικόνας, για όλη την εικόνα, για ένα καρέ ενός βίντεο κ.ά. Το δεύτερο επίπεδο αναπαριστά τις ιδιότητες του εκάστοτε είδους τμήματος πολυμέσου και αφορά ιδιότητες για χωρικά, χρονικά και χωροχρονικά τμήματα, όπως είναι για παράδειγμα συντεταγμένες περικλείοντος πλαισίου, δυαδικές μάσκες περιοχής, χρονική σφραγίδα ενός καρέ, κ.ά. Το επίπεδο οπτικής περιγραφής δημιουργεί συνδέσεις με την οντολογία οπτικών περιγραφών VDO και συνδέει τα τμήματα με περιγραφείς όπως για παράδειγμα ο περιγραφέας ομοιογενούς υφής. Η ύπαρξη αυτού του επιπέδου μεταδεδομένων εξυπηρετεί περιπτώσεις αναζήτησης με βάση μία εικόνα παράδειγμα ή ένα σκίτσο, τακτική αρκετά διαδεδομένη σε συστήματα αναζήτησης με βάση το περιεχόμενο.

Τέλος, το επίπεδο των σημασιολογικών μεταδεδομένων αποτελεί το πιο σημαντικό στην προτεινόμενη προσέγγιση περιγραφής πολυμέσων. Το επίπεδο αυτό των μεταδεδομένων ουσιαστικά συνδέει τα τμήματα με τις σημασιολογικές έννοιες που εξήχθησαν κατά τη διάρκεια της ανάλυσης της εικόνας. Οι έννοιες προέρχονται από τις οντολογίες θεματικών πεδίων και είναι αυτές που καθορίζουν το σημασιολογικό λεξιλόγιο που μπορεί να χρησιμοποιηθεί στην περιγραφή του περιεχομένου. Για παράδειγμα, μπορούμε να έχουμε την περιγραφή ότι μία εικόνα απεικονίζει ένα ορεινό τοπίο και επιπλέον ένα συγκεκριμένο τμήμα της εικόνας απεικονίζει ένα ελάφι.

Αμέσως από κάτω φαίνεται ένα απόσπασμα σημασιολογικών μεταδεδομένων για μία εικόνα από το θεματικό πεδίο “παραλία” (σχήμα 3.4):

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ns1="http://www.acemedia.org/ontologies/SCHEMA#"
  xmlns:ns2="http://www.acemedia.org/ontologies/INSTANCES#">
  <rdf:Description rdf:about="ns2:image1_segment3">
    <rdf:type rdf:resource="ns1:StillRegion"/>
    <ns1:depicts rdf:resource="ns2:Sea"/>
    <ns1:degree rdf:resource="ns2:0.214404"/>
    <ns1:depicts rdf:resource="ns2:Sand"/>
    <ns1:degree rdf:resource="ns2:0.798639"/>
  </rdf:Description>
</rdf:RDF>
```



Σχήμα 3.4: Αποτελέσματα σε εικόνες από το θεματικό πεδίο “παραλία”. (α) Εικόνα εισόδου, (β) αρχική κατάτμηση και (γ) ταξινόμηση περιοχών με ταίριασμα πρότυπων στιγμιοτύπων.

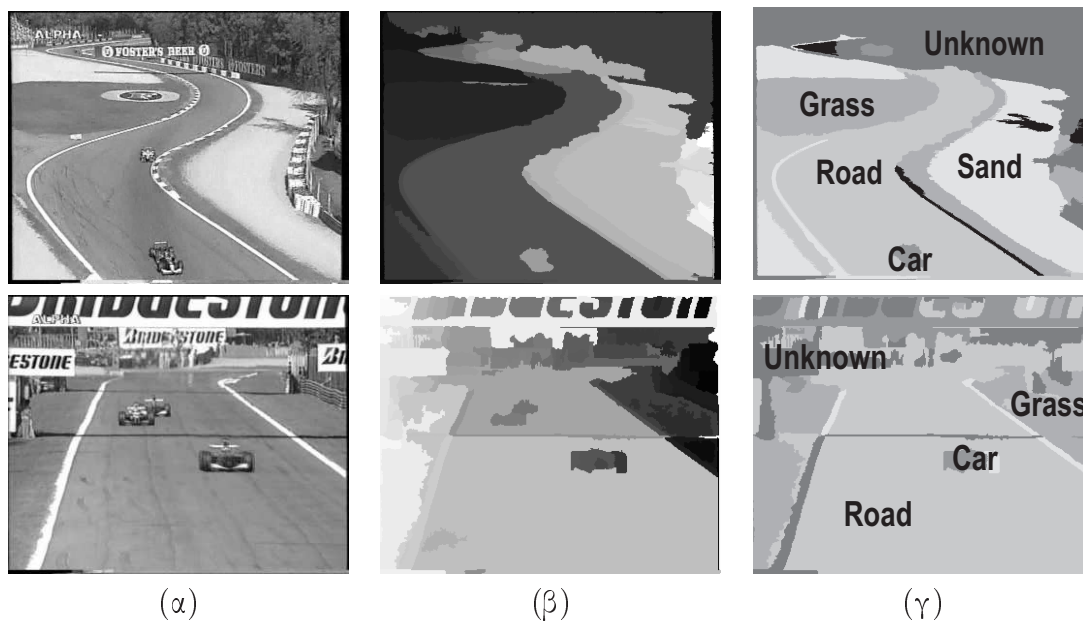
```

<ns1:aboveOf rdf:resource="ns2:image1_segment7"/>
<ns1:rightOf rdf:resource="ns2:image1_segment2"/>
<ns1:aboveAll rdf:resource="ns2:false"/>
<ns1:belowAll rdf:resource="ns2:false"/>
</rdf:Description>
</rdf:RDF>

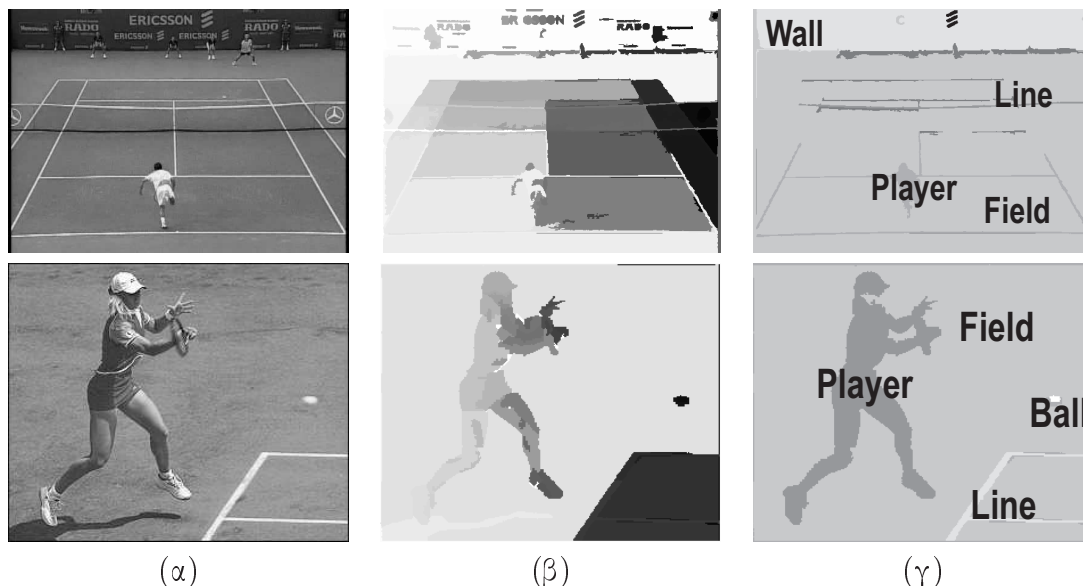
```

Στο μικρό αυτό κομμάτι των παραγόμενων μεταδεδομένων της εικόνας μπορούμε να κάνουμε τις εξής παρατηρήσεις: Καταρχήν ο ονοματοχώρος (namespace) *ns1* χρησιμοποιείται στο URI για να δηλωθεί το είδος του πολυμεσικού τμήματος που θα περιγραφεί, δηλαδή ότι το στιγμιότυπο "*ns2:image1_segment3*" είναι τύπου "*ns1:StillRegion*". Στη συνέχεια το συγκεκριμένο στιγμιότυπο περιγράφεται σημασιολογικά από τις έννοιες ΘΑΛΑΣΣΑ και ΑΜΜΟΣ και μάλιστα με έναν βαθμό εμπιστοσύνης. Τέλος στο παραπάνω απόσπασμα φαίνεται και ο τρόπος περιγραφής των χωρικών σχέσεων μεταξύ των τμημάτων που απαρτίζουν την εικόνα *image1*. Πιο συγκεκριμένα, έχουμε ότι το στιγμιότυπο "*ns2:image1_segment3*" είναι ΠΑΝΩΑΠΟ το στιγμιότυπο "*ns2:image1_segment7*", ΔΕΞΙΑΑΠΟ το στιγμιότυπο "*ns2:image1_segment2*" και δεν ικανοποιεί τα κριτήρια για να είναι η περιοχή ΠΑΝΩΑΠΟΟΛΕΣ ή ΚΑΤΩΑΠΟΟΛΕΣ τις περιοχές της εικόνας.

□



Σχήμα 3.5: Αποτελέσματα σε εικόνες από το θεματικό πεδίο “φόρμουλα 1”. (α) Εικόνα εισόδου, (β) αρχική κατάτμηση και (γ) ταξινόμηση περιοχών με ταίριασμα πρότυπων στιγμιοτύπων.



Σχήμα 3.6: Αποτελέσματα σε εικόνες από το θεματικό πεδίο “τένις”. (α) Εικόνα εισόδου, (β) αρχική κατάτμηση και (γ) ταξινόμηση περιοχών με ταίριασμα πρότυπων στιγμιοτύπων.

Κεφάλαιο 4

Σημασιολογική Κατάτμηση Εικόνων και Βίντεο

4.1 Κατάτμηση Εικόνων

Η όραση υπολογιστών ασχολείται συστηματικά με τη θεωρία και την εφαρμογή αλγορίθμων για την αυτοματοποίηση διαδικασιών οπτικής αντίληψης. Χοντρικά, το αντικείμενο της έρευνας διαχωρίζεται σε διαδικασίες και μηχανισμούς χαμηλού επιπέδου, όπως για παράδειγμα η αφαίρεση θορύβου, ο εντοπισμός και εξομάλυνση ή όξυνση ακμών, η οπτική περιγραφή εικόνων και περιοχών ενδιαφέροντος, η κατάτμηση σε ομογενείς περιοχές και σε διαδικασίες υψηλού επιπέδου όπως είναι η ταξινόμηση εικόνων, ο εντοπισμός και η αναγνώριση αντικειμένων και γενικότερα η κατανόηση περιεχομένου [121, 46]. Είναι βέβαια εμφανής η συσχέτιση και η επικάλυψη των δύο αυτών επιπέδων, και πιο συγκεκριμένα για την κατάτμηση, η οποία αποτελεί ένα πρώτο βήμα για την εξαγωγή σημασιολογικών περιοχών που οδηγούν στην κατανόηση της εικόνας.

Η αυτόματη κατάτμηση εικόνων είναι ένα από τα θεμελιώδη προβλήματα στην επεξεργασία εικόνας και την όραση υπολογιστών και είναι μία εξόχως απαιτητική και δύσκολη εργασία. Ο αντικειμενικός σκοπός της κατάτμησης είναι ο διαχωρισμός της εικόνας σε σημαίνουσες περιοχές οι οποίες να έχουν ισχυρή συσχέτιση με αντικείμενα τα οποία περιέχονται εξ' ολοκλήρου ή εν μέρει στην εικόνα [121]. Η δυσκολία του προβλήματος γίνεται περισσότερο έκδηλη όταν προσπαθήσουμε να ορίσουμε την έννοια “αντικείμενο”. Λαμβάνοντας υπόψιν την πολυπλοκότητα των εικόνων, γίνονται δυσδιάκριτα τα όρια των αντικειμένων, όπως μπορεί να φανεί από τα σχήματα 4.1 και 4.2. Στα εν λόγω δύο σχήματα έχουμε τα αποτελέσματα της χειρωνακτικής κατάτμησης δύο εικόνων από διαφορετικούς ανθρώπους, όπου μπορούμε να παρατηρήσουμε το πόσο υποκειμενική είναι η διαδικασία κατάτμησης μίας εικόνας. Οι εικόνες προέρχονται από τη βάση δεδομένων και αξιολόγησης κατάτμησης εικόνων του Berkeley [73] και οι οδηγίες που δόθηκαν στους χρήστες ήταν να διαιρέσουν την κάθε εικόνα σε τμήματα τα οποία να αναπαριστούν “πράγματα” ή “μέρη πραγμάτων” στην σκηνή, όμοιας ή ίδιας, όσο είναι αυτό δυνατόν, σημασίας. Κάθε ένας κατάτμησε την εικόνα σε διαφορετικό βαθμό τμημάτων, ανάλογα με το βαθμό λεπτομέρειας τον οποίο θεώρησε ότι αντιπροσωπεύει καλύτερα την σημασιολογική ερμηνεία της σκηνής.

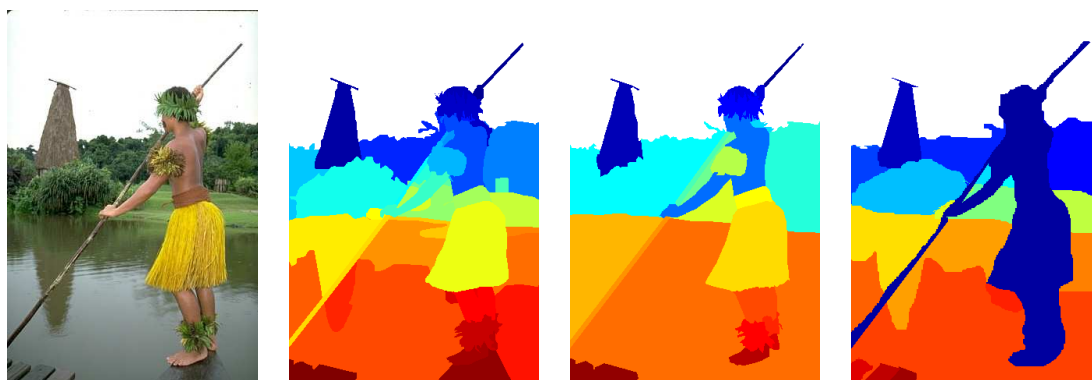
Εφόσον δε μπορούμε να ορίσουμε τι ακριβώς εννοούμε ως “σημαίνουσα” περιοχή, τότε είναι σαφές ότι προσπαθούμε να λύσουμε ένα πρόβλημα που δεν έχει διατυπωθεί επακριβώς και δεν γνωρίζουμε ποιος είναι ο στόχος της κατάτμησης. Αν σκοπός

μας είναι η πλήρης κατάτμηση μίας εικόνας, τότε το αποτέλεσμα είναι ένα σύνολο ασύνδετων περιοχών οι οποίες αντιστοιχούν μία προς μία με αντικείμενα ή, γενικότερα, έννοιες που υπάρχουν στην αρχική εικόνα. Μία πιο χαλαρή απαίτηση είναι αυτή για μερική κατάτμηση, όπου για παράδειγμα επιθυμούμε να διαχωρίσουμε ένα ή έστω κάποια αντικείμενα από το παρασκήνιο. Μία άλλη πιο χαλαρή απαίτηση είναι να ομαδοποιήσουμε περιοχές με βάση κάποιο κριτήριο ομοιογένειας όχι απαραίτητα ασύνδετες μεταξύ τους, επιτρέποντας ακόμα και επικαλύψεις περιοχών.

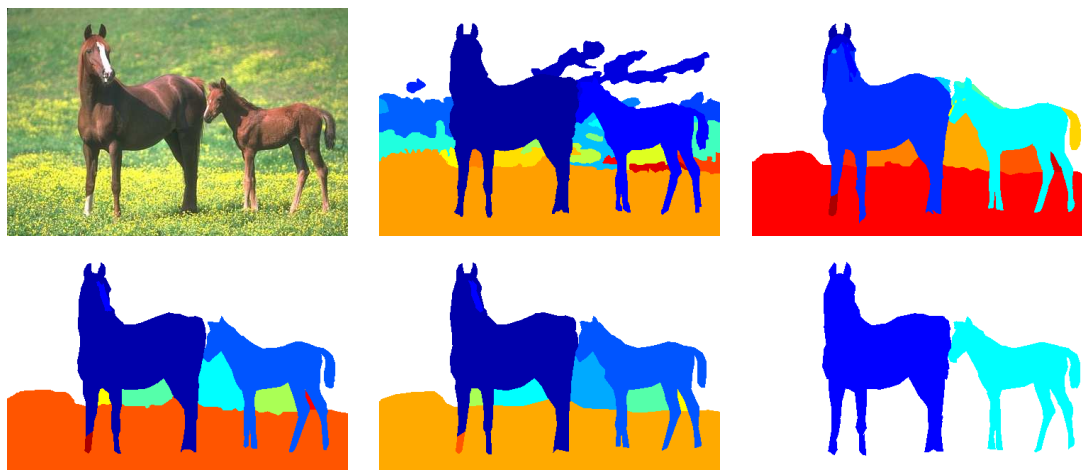
Παρά την ελλιπή διατύπωση του προβλήματος της κατάτμησης, μία μεγάλη γκάμα εφαρμογών και συστημάτων όπως η κωδικοποίηση εικόνων, η αναγνώριση αντικειμένων, η τεκμηρίωση και δεικτοδότηση, μεταξύ άλλων, χρησιμοποιούν σε κάποιο στάδιο κάποιον αλγόριθμο κατάτμησης και η απόδοσή τους εξαρτάται σημαντικά από αυτόν. Είναι κοινώς αποδεκτό ότι παρά την πολυετή έρευνα στο αντικείμενο, η οποία έχει οδηγήσει σε μία πληθώρα τεχνικών και αλγορίθμων βασισμένων σε διαφορετικές προσεγγίσεις [105, 93, 26], υπάρχουν ορισμένα ανυπέρβλητα προβλήματα, όπως είναι για παράδειγμα, η μερική ή ολική επικάλυψη (occlusion), η σπονδυλωτή κίνηση (non-rigid motion), η υπερκατάτμηση (oversegmentation), τα οποία μειώνουν την επίδοση, την ακρίβεια, την ευστάθεια και κατά συνέπεια την αξιοπιστία των προαναφερθεισών εφαρμογών. Σε αντιδιαστολή με άλλα αντικείμενα της όρασης υπολογιστών τα οποία είναι καλά θεμελιωμένα, όπως π.χ. της αναγνώρισης και της ταξινόμησης, η κατάτμηση είναι ένα ασθενώς ορισμένο (ill-posed) πρόβλημα, καθώς δεν υφίσταται μοναδική γενική λύση, αντικειμενικά και καθολικά αποδεκτή ενώ και η αξιολόγησή της επίδοσης του εκάστοτε αλγορίθμου κατάτμησης παραμένει μία διαδικασία κατεξοχήν υποκειμενική. Στην κατάτμηση, είναι πολύ δύσκολο να ποσοτικοποιηθεί η επίδοση μίας μεθόδου, ούτως ώστε να είναι εφικτή η άμεση σύγκριση δύο αλγορίθμων μετρώντας την ίδια ποσότητα σφάλματος. Προκειμένου να μετατραπεί το πρόβλημα της κατάτμησης σε καλά ορισμένο (well-posed) και να καταστεί δυνατή η επίλυσή του, χρειάζεται να επαναδιατυπωθεί και να γίνουν ορισμένες υποθέσεις ώστε να περιοριστεί το πλήθος των αποδεκτών λύσεων.

4.1.1 Ταξινομία Αλγορίθμων Κατάτμησης

Οι τεχνικές κατάτμησης εικόνων που έχουν προταθεί στη βιβλιογραφία μπορούν να ταξινομηθούν με βάση διάφορα κριτήρια. Πολύ συχνά ο διαχωρισμός γίνεται σε δύο ομάδες: (i) σε τεχνικές με βάση τις περιοχές, που βασίζονται σε κριτήρια ομοιογένειας χαρακτηριστικών χωρικά ομαδοποιημένων εικονοστοιχείων και (ii) σε τεχνικές



Σχήμα 4.1: Χειρωνακτική κατάτμηση εικόνας από τη βάση εικόνων του Berkeley [73]. Από αριστερά προς τα δεξιά έχουμε πιο χονδροειδή κατάτμηση με λιγότερες περιοχές.



Σχήμα 4.2: Κατάτμηση εικόνας από διαφορετικούς χρήστες (2ο παράδειγμα από τη βάση εικόνων του Berkeley).

με βάση το περίγραμμα, όπου εντοπίζονται τα όρια των αντικειμένων (π.χ. με χρήση της χωρικής παραγώγου - gradient). Πιο λεπτομερής διαχωρισμός επιτυγχάνεται αν ως κριτήριο ταξινόμησης χρησιμοποιηθεί η μαθηματική μεθοδολογία πάνω στην οποία βασίζεται ένας αλγόριθμος κατάτμησης [70]. Με αυτή τη λογική προκύπτουν οι εξής μέθοδοι: (i) μεταβολικές (variational), (ii) στατιστικές, (iii) βασιζόμενες σε γράφους και (iv) μορφολογικές. Ένα άλλο σημαντικό κριτήριο για να ταξινομήσουμε μεθόδους κατάτμησης είναι το είδος της πληροφορίας που χρησιμοποιείται για την ομαδοποίηση, σύμφωνα με το οποίο υπάρχουν χοντρικά δύο κατηγορίες: (i) αλγόριθμοι οδηγούμενοι από τα δεδομένα ή αλλιώς εκ των κάτω προσέγγιση (bottom-up) και (ii) αλγόριθμοι οδηγούμενοι από το μοντέλο ή αλλιώς άνωθεν προσέγγιση (top-down). Οι τεχνικές της πρώτης κατηγορίας προσπαθούν να εξαγάγουν σημαντικές οντότητες (ιδεατά αντικείμενα) αποκλειστικά από την επεξεργασία οπτικών χαρακτηριστικών όπως είναι το χρώμα ή η ομοιογένεια υφής και βασίζονται σε επαναληπτικές διαδικασίες συγχώνευσης εικονοστοιχείων και περιοχών. Οι τεχνικές της δεύτερης κατηγορίας προσπαθούν να κατατμήσουν την εικόνα σε αντικείμενα των οποίων τα οπτικά χαρακτηριστικά ταιριάζουν με αυτά των μοντέλων.

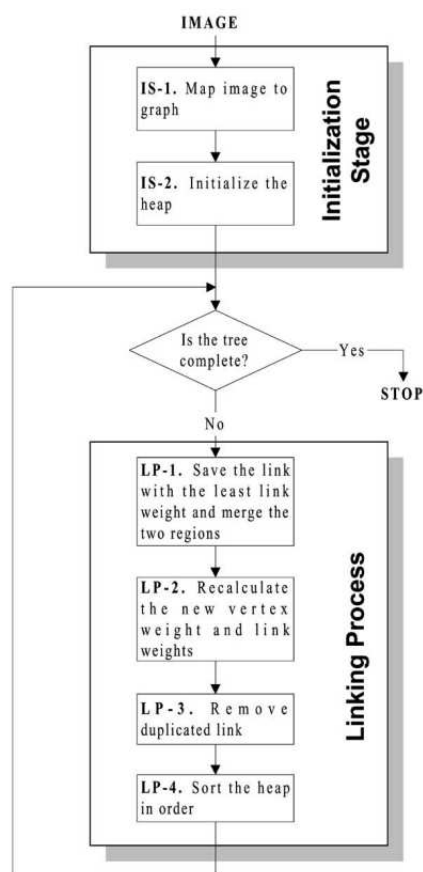
Η βιβλιογραφία στο αντικείμενο της κατάτμησης εικόνων είναι τεράστια: ενδεικτικά αναφέρουμε ορισμένες διαδεδομένες μεθοδολογίες κατάτμησης όπως η κατωφλίωση ιστογράμματος [92], με βάση τις ακμές [97, 99], η συσταδοποίηση (clustering) [16, 57, 25, 27], η μαθηματική μορφολογία [139, 77], γραφοθεωρητικές τεχνικές [82, 104, 109, 60, 39], στατιστικές τεχνικές [134] και τα νευρωνικά δίκτυα [94] μεταξύ πολλών άλλων.

Πιο πρόσφατα, ένα μεγάλο κομμάτι της έρευνας έχει εστιάσει στο συνδυασμό χαρακτηριστικών τοπικών περιοχών της εικόνας, με χαρακτηριστικά εξαγόμενα από όλη την εικόνα. Η λογική αυτής της προσέγγισης είναι ότι πληροφορία που έχει εξαχθεί από όλη την εικόνα μπορεί να βοηθήσει να ξεκαθαρίσουν ασάφειες που προέρχονται από την ανάλυση σε περιοχές της εικόνας. Στην εργασία [66] χρησιμοποιήθηκε ένα δίκτυο Bayes για να ενσωματώσει χαμηλού και μέσου επιπέδου χαρακτηριστικά με σκοπό την ταξινόμηση εικόνων εσωτερικού ή εξωτερικού χώρου. Στην εργασία [22] προτάθηκε μία πολυκλασική ταξινόμηση, παρόμοια με προσεγγίσεις ασαφούς λογικής, για να περιγράψει φυσικές εικόνες στις οποίες συνήθως περιέχονται πολλά σημασιολογικά αντικείμενα, ενώ ασαφής ταξινόμηση περιοχών χρησιμοποιήθηκε στην εργασία

[61]. Στην εργασία [88], με εκμάθηση δικτύου Bayes με ταξινομητές ως κόμβους, αντιστοιχίζεται η χαμηλού επιπέδου αναπαράσταση βίντεο με υψηλού επιπέδου έννοιες για δεικτοδότηση, φιλτράρισμα και ανάκτηση βίντεο. Επίσης, η γνώση για τη σχετική θέση μεταξύ περιοχών στην εικόνα μπορεί να χρησιμοποιηθεί για να περιοριστεί το εύρος των πιθανών λύσεων και να βελτιωθεί η επίδοση της αναγνώρισης αντικειμένων [78]. Τέλος έχει προταθεί η χρήση του αλγορίθμου EM [59] για την συνέργεια της αναγνώρισης αντικειμένων και της κατάτμησης εικόνων.

4.1.2 Αλγόριθμος Κατάτμησης RSST

Ως βάση για την κατάτμηση εικόνων στο πλαίσιο της διατριβής, χρησιμοποιήθηκε ο αλγόριθμος αναδρομικού ελάχιστου συνδετικού δένδρου (recursive shortest spanning tree - RSST) [82]. Παρά τη σχετικά μεγάλη υπολογιστική του πολυπλοκότητα, θεωρείται ένα ισχυρό εργαλείο κατάτμησης εικόνων συγκρινόμενο με άλλες τεχνικές [84] (όπως η συσταδοποίηση με βάση το χρώμα, η πυραμιδική επέκταση περιοχών, ο μορφολογικός αλγόριθμος watershed), που προτιμάται λόγω της δυνατότητας εύκολου ορισμού του κριτηρίου συγχώνευσης περιοχών.



Σχήμα 4.3: Διάγραμμα ροής του αλγορίθμου κατάτμησης RSST [82].

Το διάγραμμα ροής του αλγορίθμου φαίνεται στο σχήμα 4.3 και η λειτουργία του περιγράφεται συνοπτικά στη συνέχεια. Κατά την αρχικοποίηση του αλγορίθμου η εικόνα εισόδου I , διαστάσεων $M \times N$ χωρίζεται σε $M \times N$ περιοχές μεγέθους ενός εικονοστοιχείου και για κάθε περιοχή παράγονται σύνδεσμοι (links) με τις τέσσερις γειτονικές της περιοχές. Γειτονικές μιας περιοχής X θεωρούμε τις περιοχές εκείνες

που απέχουν ένα εικονοστοιχείο από τη X και συνδέονται μαζί της οριζόντια ή κατακόρυφα, αλλά όχι διαγώνια (4-connected). Σε κάθε σύνδεσμο ανατίθεται ένα βάρος, ίσο με την απόσταση των δύο περιοχών τις οποίες συνδέει. Η απόσταση αυτή μπορεί να ορισθεί με οποιονδήποτε τρόπο και καθορίζει ουσιαστικά την προτεραιότητα με την οποία οι περιοχές ενώνονται μεταξύ τους. Για παράδειγμα, για τους σκοπούς της κατάτμησης με βάση το χρώμα η απόσταση $d_c(X, Y)$ μεταξύ δύο γειτονικών περιοχών X και Y συνήθως ορίζεται ως το γινόμενο της Ευκλείδειας απόστασης των μέσων χρωματικών συνιστωσών των δύο περιοχών και ενός όρου που είναι συνάρτηση του μεγέθους των δύο περιοχών σε pixels και ευνοεί τη συγχώνευση των μικρών περιοχών:

$$d_c(X, Y) = \|c(X) - c(Y)\| \cdot \frac{|X| \cdot |Y|}{|X| + |Y|} \quad (4.1)$$

όπου $c(X)$ είναι ένα διάνυσμα διαστάσεων 3×1 που περιέχει τις τρεις μέσες χρωματικές συνιστώσες της περιοχής X και $|X|$ είναι το μέγεθός της σε pixels. Στη συνέχεια όλοι οι σύνδεσμοι ταξινομούνται κατά αύξουσα σειρά απόστασης, έτσι ώστε ο πρώτος σύνδεσμος με τη μικρότερη απόσταση να αντιστοιχεί στις πιο “κοντινές” περιοχές.

Κατά την επαναληπτική φάση του RSST πραγματοποιείται αναδρομική συγχώνευση επιλέγοντας κάθε φορά τις δύο γειτονικές περιοχές με τη μικρότερη απόσταση και εφαρμόζοντας τις ακόλουθες ενέργειες σε κάθε επανάληψη:

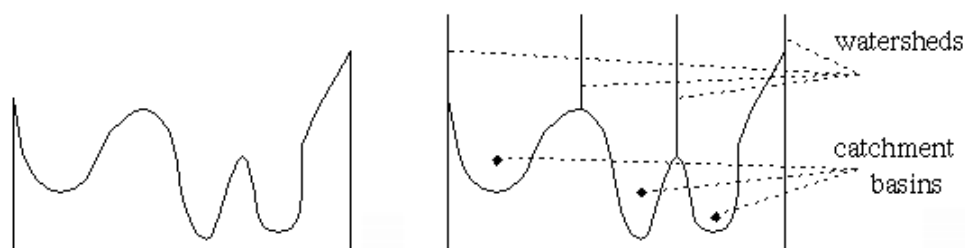
1. Συγχωνεύονται οι δύο πλησιέστερες γειτονικές περιοχές και επαναυπολογίζονται οι ιδιότητες της νέας περιοχής (π.χ. οπτικοί περιγραφείς για το χρώμα την υφή, το μέγεθος).
2. Επαναυπολογίζονται και ταξινομούνται τα βάρη των συνδέσμων της νέας περιοχής με όλες τις γειτονικές της.
3. Αφαιρούνται οι τυχόν διπλοί σύνδεσμοι που προκύπτουν από τη συγχώνευση.

Η επαναληπτική διαδικασία τερματίζεται όταν είτε ο συνολικός αριθμός περιοχών είτε η ελάχιστη απόσταση γίνει ίση με ένα προκαθορισμένο κατώφλι. Το κατώφλι απόστασης είναι συνήθως προτιμότερο διότι το πλήθος των τμημάτων που προκύπτει από την κατάτμηση εξαρτάται από τα αντικείμενα που υπάρχουν στην εικόνα. Η υπολογιστική πολυπλοκότητα του RSST καθορίζεται σε μεγάλο βαθμό από την επιλογή του αλγορίθμου ταξινόμησης των βαρών, ο οποίος αποτελεί το βασικό παράγοντα καθυστέρησης.

4.1.3 Αλγόριθμος Κατάτμησης Watershed

Ο μετασχηματισμός watershed οφείλει το όνομά του στην τοπολογία, όπου μελετώνται το σχήμα και τα γεωγραφικά χαρακτηριστικά (βουνά, ποτάμια, κλπ.) της επιφάνειας της γης. Πρώτοι οι Digabel και Lantuejoul [35] πρότειναν το μετασχηματισμό watershed για την κατάτμηση εικόνων και έκτοτε μελετήθηκε και επεκτάθηκε [77, 139, 29, 120] και αποτελεί μία κλασική τεχνική που υπάγεται στο πεδίο της μαθηματικής μορφολογίας.

Η βασική ιδέα είναι να θεωρηθεί η εικόνα (ή συνήθως μία συνάρτηση ύψους όπως π.χ. η κλίση της φωτεινότητας - gradient) ως μία τοπολογική επιφάνεια: η τιμή του εικονοστοιχείου αντιστοιχεί στο ύψος του συγκεκριμένου σημείου. Στο σχήμα 4.4 φαίνεται το μονοδιάστατο προφίλ μίας εικόνας, όπου έχουν εντοπιστεί τα τοπικά



Σχήμα 4.4: Μονοδιάστατο παράδειγμα κατάτμησης watershed. Τα τοπικά ελάχιστα δημιουργούν κοιλάδες περισυλλογής, ενώ τα τοπικά μέγιστα ορίζουν τις υδατογραμμές [121].

ελάχιστα και τα τοπικά μέγιστα. Το σύνολο των εικονοστοιχείων που συνδέονται με ένα τοπικό ελάχιστο δημιουργεί μία κοιλάδα περισυλλογής (catchment basin), ενώ τα εικονοστοιχεία που βρίσκονται σε τοπικά μέγιστα (ή ισοδύναμα, δεν ανήκουν σε καμία κοιλάδα περισυλλογής) ορίζουν τις υδατογραμμές (watershed lines), οι οποίες και καθορίζουν το περίγραμμα της κάθε περιοχής μετά το πέρας της κατάτμησης.

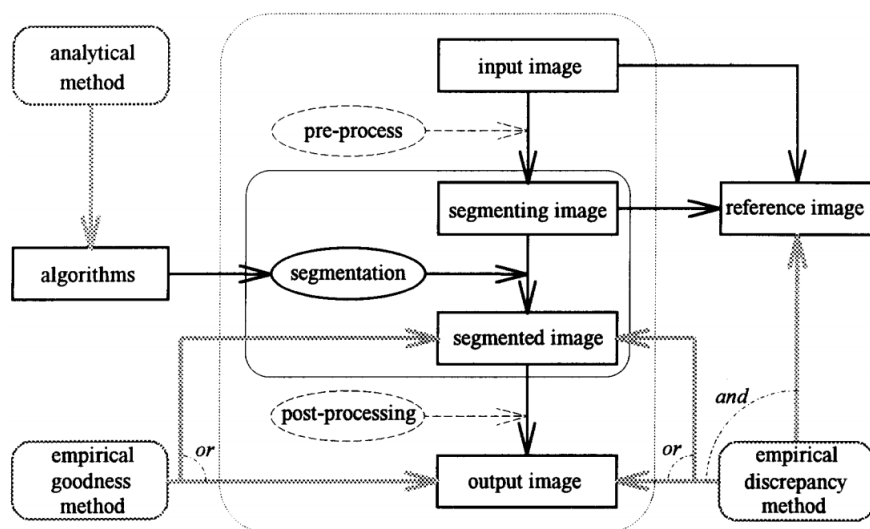
Η διαδικασία πλημμυρισμού ξεκινάει από τα τοπικά ελάχιστα της τοπογραφικής επιφάνειας της εικόνας και το ύψος του νερού ανεβαίνει έως ότου δύο κοιλάδες (ή λίμνες) συναντηθούν, οπότε και σχηματίζεται ένα φράγμα το οποίο αποτελεί τη διαχωριστική υδατογραμμή των δύο αυτών περιοχών. Ο πλημμυρισμός συνεχίζεται έως ότου να καλυφθεί όλη η επιφάνεια από νερό, ενώ παράλληλα δημιουργούνται οι υδατογραμμές που θα ορίσουν την τελική κατάτμηση της εικόνας.

Ένα βασικό μειονέκτημα του watershed αποτελεί ο μεγάλος αριθμός τοπικών ελαχίστων που οδηγεί σε μεγάλο αριθμό τελικών περιοχών, με άλλα λόγια σε υπερκατάτμηση της εικόνας. Για να αντιμετωπιστεί αυτό το πρόβλημα, γίνεται χρήση σηματοδευτών (markers), οι οποίοι καθορίζουν ποιες περιοχές θα πλημμυρίσουν, περιορίζοντας τον αριθμό των τελικών περιοχών στον αριθμό των σηματοδευτών. Η διαδικασία επιλογής σηματοδευτών είναι δύσκολη και ιδιαίτερα κρίσιμη για το τελικό αποτέλεσμα και εξαρτάται και από την εκάστοτε εφαρμογή.

4.1.4 Αξιολόγηση Αλγορίθμων Κατάτμησης

Είναι παράδοξο το γεγονός ότι παρά την πολυετή έρευνα στο αντικείμενο της κατάτμησης εικόνων, συγκριτικά έχει αφιερωθεί πολύ λιγότερη προσπάθεια για ένα από τα βασικότερα στοιχεία οποιασδήποτε έρευνας, αυτό της αξιολόγησης. Μάλλον έχει υποτιμηθεί από την συγκεκριμένη επιστημονική κοινότητα η αναγκαιότητα και η χρησιμότητα για αντικειμενικά ή υποκειμενικά κριτήρια αξιολόγησης, διαθέσιμα σύνολα δεδομένων δοκιμών, τυπικές και κοινώς αποδεκτές μετρικές, επιστημονικές συγκρίσεις και αποτελέσματα.

Πρώιμες εργασίες στο αντικείμενο εμφανίστηκαν στα τέλη της δεκαετίας του 70, εστιάζοντας στην αξιολόγηση αλγορίθμων εντοπισμού ακμών, οι οποίοι αποτελούν ένα πολύ μικρό υποσύνολο της κατάτμησης εικόνων (αν και δεν υπάρχει κάποιος κοινώς αποδεκτός ορισμός του τι είναι κατάτμηση εικόνας). Η πρώτη σημαντική εργασία που έκανε επισκόπηση της περιοχής της αξιολόγησης τεχνικών κατάτμησης



Σχήμα 4.5: Η γενική κατάταξη κατά τον Zhang για τις μεθοδολογίες αξιολόγησης αλγορίθμων κατάτμησης εικόνων [146].

εικόνων ήταν από τον Zhang [146], όπου παρουσίασε μία πλήρη κατηγοριοποίηση μεθοδολογιών αξιολόγησης κατάτμησης. Πιο συγκεκριμένα, ταξινόμησε τις μεθόδους σε τρεις γενικές κατηγορίες (σχήμα 4.5):

- Αναλυτικές, οι οποίες προσεγγίζουν θεωρητικά το πρόβλημα, όπως για παράδειγμα η μελέτη της πολυπλοκότητας
- Εμπειρικής συμβατότητας (goodness), που χρησιμοποιούν μετρικές βασισμένες στα αποτελέσματα αλγορίθμων κατάτμησης εικόνων.
- Εμπειρικής ανακολουθίας (discrepancy), που μετρούν μία απόσταση από μία ιδεατή κατάτμηση από το σύνολο δεδομένων αλήθειας (ground truth).

Ο Zhang εξέτασε το πρόβλημα από μία σχετικά υψηλού επιπέδου οπτική, χωρίς να δώσει πολλές τεχνικές λεπτομέρειες. Παρόλα αυτά, θεμελίωσε τις επιθυμητές ιδιότητες που πρέπει να έχει μία μεθοδολογία αξιολόγησης: να είναι α) ανεξάρτητη εφαρμογής, β) ποσοτική και αντικειμενική και γ) χαμηλής πολυπλοκότητας.

Μία ακόμα ενδεδειγμένη επισκόπηση παρουσιάστηκε από τους Hoover et al. [50], όπου μερικές ερευνητικές ομάδες παρουσίασαν μία κοινή πλατφόρμα για τη δοκιμή και την αξιολόγηση διάφορων αλγορίθμων κατάτμησης και μετρικών αξιολόγησης. Αν και οι αλγόριθμοι κατάτμησης καθώς και το σύνολο δεδομένων θεωρούνται πλέον ξεπερασμένα, το άρθρο αυτό παρουσίασε έναν καλό ορισμό του προβλήματος, μία πλήρη μεθοδολογία αξιολόγησης (σε αντίθεση με τον Zhang [146]) και διέθεσε τόσο τους αλγορίθμους όσο και το σύνολο των δεδομένων στην ερευνητική κοινότητα για περαιτέρω αξιολόγηση και κρίση. Η κύρια συνεισφορά ήταν η χρήση μετρικής βασισμένης στο σύνολο δεδομένων αλήθειας με χρήση επικαλυπτόμενων περιοχών και διαφόρων κριτηρίων για υπέρ- και υπό-κατάτμηση και θόρυβο. Χρήση συνόλου δεδομένων αλήθειας έκαναν και οι Huang και Dom [52] που πρότειναν τρεις μεθόδους αξιολόγησης: μία βασισμένη σε υπολογισμό παραμέτρων, μία δεύτερη βασισμένη σε χαρακτηριστικά ακμών και περιγραμμάτων και μία τρίτη με τη μεγαλύτερη επίδραση, βασισμένη σε υπολογισμό απόστασης Hamming μεταξύ των επικαλυπτόμενων περιοχών.

Πιο πρόσφατη είναι η δουλειά και μετέπειτα η διατριβή του Martin [73, 72] η οποία αποτέλεσε τομή και σημείο αναφοράς για μελλοντικές εργασίες καθώς εδραίωσε την πλατφόρμα συγκριτικής αξιολόγησης Berkeley, η οποία περιλαμβάνει 300 εικόνες με τις μάσκες κατάτμησης από πολλούς χρήστες. Παρουσίασαν δύο μετρικές σφάλματος κατάτμησης, το τοπικό σφάλμα συνέπειας (Local Consistency Error - LCE) και το καθολικό σφάλμα συνέπειας (Global Consistency Error - GCE), τα οποία υπολογίζουν με μη συμμετρικό τρόπο την επικάλυψη των περιοχών που δημιουργούνται από την αυτόματη κατάτμηση με τις περιοχές που έχουν χαράξει οι χρήστες.

Πιο πρόσφατα, έχουν επικρατήσει οι τεχνικές αξιολόγησης που ανήκουν στην κατηγορία της εμπειρικής ανακολουθίας και έχουν προταθεί ιδέες είτε πιο γενικές είτε εστιασμένες ανά πεδίο εφαρμογής. Ένας δείκτης που υπολογίζει την ακρίβεια της κατάτμησης και λαμβάνει υπόψιν του ορισμένες επιπλέον απαιτήσεις (για να επιλύσει γνωστά προβλήματα συναφών τεχνικών) όπως να μην γίνεται υπόθεση για τον αριθμό των περιοχών ή να υπάρχει έλεγχος για εκφυλιστικές ακραίες περιπτώσεις [135]. Επίσης υπάρχει η άποψη ότι δεν είναι απαραίτητη η αξιολόγηση της κατάτμησης μίας εικόνας, αλλά απαιτείται καλή επίδοση στην εξαγωγή των σημαντικών περιοχών που αναπαριστούν κάποιο αντικείμενο [43, 44]. Τέλος έχουν γίνει προσπάθειες για ενσωμάτωση τεχνικών αξιολόγησης κατάτμησης σε αυτόματα και ημιαυτόματα συστήματα κατάτμησης. Μία τέτοια προσπάθεια παρουσιάστηκε στις εργασίες [74], όπου μεταξύ άλλων χρησιμοποιήθηκε η βάση δεδομένων του Berkeley [73], και κατέληξαν ότι η αυτόματη εξαγωγή του φόντου σε συνδυασμό με μία τεχνική κατάτμησης μπορεί να αντικαταστήσει τη χειρωνακτική εργασία της κατασκευής του συνόλου δεδομένων αλήθειας χωρίς μεγάλο κόστος στην ποιότητα αυτών [45].

4.2 Σημασιολογική Επέκταση Περιοχών

Παρά τη μεγάλη ερευνητική προσπάθεια, η ανθρώπινη οπτική αντίληψη υπερέρχει ξεκάθαρα των πιο σύγχρονων αλγορίθμων κατάτμησης. Ένας από τους κυριότερους λόγους για αυτό είναι ότι η όραση του ανθρώπου βασίζεται σε ένα μεγάλο ποσοστό στην τεράστια πρότερη (a priori) γνώση που έχει συσσωρεύσει μέσα από τα βιώματά του ο κάθε άνθρωπος και ο τρόπος που μπορεί να τη διαχειριστεί [71]. Η γνώση αυτή έχει να κάνει με τη σημασιολογική ερμηνεία των αντικειμένων όπως αυτά εμφανίζονται μέσα σε μία εικόνα (σχετική θέση, χρώμα, σχήμα, κλπ.) και βοηθάει να συμπληρωθούν τα κενά που αναπόφευκτα υπάρχουν από την 2Δ απεικόνιση της 3Δ πραγματικότητας. Η διαπίστωση αυτή οδήγησε στην εισαγωγή νέων ιδεών από το χώρο της τεχνητής νοημοσύνης, ώστε να βρεθεί αποδοτικός τρόπος εξαγωγής, αναπαράστασης, διαχείρισης και αξιοποίησης της γνώσης του φυσικού κόσμου σε προβλήματα κατάτμησης εικόνων. Προφανώς, εξαιτίας της γενικότητας του προβλήματος, ο αριθμός των υποθέσεων για μία τέτοια βάση γνώσης είναι τεράστιος και η επιλογή της εξειδικευμένης γνώσης για την εκάστοτε σκηνή, αποτελεί περίπλοκο πρόβλημα από μόνη της.

Είναι διαισθητικά φανερό ότι αν ενσωματώσουμε περισσότερη πληροφορία, η οποία να προσδίδει μία υποθετική ερμηνεία της σκηνής, θα έχουμε ευεργετική συνέπεια στη διαδικασία της κατάτμησης και πιο συγκεκριμένα στην απόφαση συγχώνευσης ή όχι δύο περιοχών. Είναι επίσης διαισθητικά σαφές ότι το (οπτικό) εννοιολογικό πλαίσιο (visual context) όπως προκύπτει από την ολική οπτική και ερμηνεία της σκηνής, δίνει αρκετούς περιορισμούς μειώνοντας έτσι αρκετά το μέγεθος των υποθέσεων που μπορούν να εξαχθούν από τη βάση γνώσης, βελτιώνοντας τις επιδόσεις τεχνικών αναγνώρισης αντικειμένων [129, 131] και καθιστώντας εφικτή και αποδοτική την

υλοποίηση σημασιολογικά ενήμερων (semantically aware) αλγορίθμων κατάτμησης εικόνων. Στόχος μας είναι να βελτιώσουμε τα αποτελέσματα της κατάτμησης εικόνας ενσωματώνοντας γνώση που εξάγεται από ανίχνευση αντικειμένων. Η συνήθης πρακτική κατανόησης μίας σκηνής είναι η σειριακή εφαρμογή των διαδικασιών της κατάτμησης και ταξινόμησης περιοχών, όπου η διαδικασία της κατάτμησης δημιουργεί ομογενείς περιοχές οι οποίες έπειτα αντιστοιχούνται σε πιθανά αντικείμενα. Η δική μας προσέγγιση ενσωματώνει την αναγνώριση μέσα στον αλγόριθμο της κατάτμησης, μιμούμενη εν μέρει τον τρόπο που λειτουργεί η ανθρώπινη όραση, καθώς έχει διαπιστωθεί ότι ο εγκέφαλος εντοπίζει γνωστά σε αυτόν αντικείμενα και μετά μπορεί να τα διαχωρίσει ευκολότερα από το φόντο, χωρίς να εξετάσει λεπτομερή χαρακτηριστικά [4].

Σχετικές εργασίες που αναδεικνύουν τη χρησιμότητα της αναγνώρισης αντικειμένων και της σημασιολογικής ερμηνείας μίας σκηνής ως έμφυτο μέρος της διαδικασίας κατάτμησης της εικόνας έχουν εμφανιστεί από τα πρώτα στάδια της σύζευξης των κλάδων της τεχνητής νοημοσύνης με την όραση υπολογιστών. Από τις πρώτες εργασίες υπήρξε αυτή των Feldman και Yakimovsky [38] η οποία πρότεινε τη συγχώνευση γειτονικών περιοχών με κοινή ερμηνεία και τη χρήση ευρετικών μεθόδων για τον τερματισμό της διαδικασίας επέκτασης περιοχών. Πέρασαν αρκετά χρόνια έως ότου να επανεμφανιστεί το ζήτημα του κατά πόσο η κατάτμηση προηγείται ή μπορεί να ιδωθεί ως μέρος της διαδικασίας αναγνώρισης αντικειμένων [98] και ιδιαίτερα η εργασία των Borenstein και Ullman [21, 20] έδωσε νέα ώθηση στη χρήση πρότερης γνώσης για την κατάτμηση εικόνων. Πιο συγκεκριμένα, πρότειναν τη χρήση γνώσης για το σχήμα ορισμένων αντικειμένων η οποία οδηγεί την κάτωθεν (bottom-up) κατάτμηση. Την ταυτόχρονη αναγνώριση αντικειμένων και κατάτμηση πρότειναν και οι Yu et al. [145] χρησιμοποιώντας ένα μηχανισμό φασματικής διχοτόμησης γράφου για τον υπολογισμό της βέλτιστης κατάτμησης της εικόνας. Οι Leibe και Schiele [62] υποστηρίζουν και αυτοί με τη σειρά τους ότι δύο διαδικασίες της αναγνώρισης και της κατάτμησης είναι περιπλεγμένες (interleaved) και χρησιμοποιούν μία πιθανολογική διατύπωση και υποθέσεις για πιθανά αναγνωρισμένα αντικείμενα για να βελτιώσουν την τελική κατάτμηση της εικόνας. Τέλος, οι Shotton et al. [111] πρότειναν νέα χαρακτηριστικά τα οποία περιγράφουν το σχήμα και την υφή μίας περιοχής και με χρήση αυτών ως εξαρτημένα τυχαία πεδία (conditional random fields - CRFs) ταξινομούν και εντοπίζουν αντικείμενα από 21 διαφορετικές κλάσεις. Στην πιο πρόσφατη σχετική δουλειά των Shotton et al. [110], αντικατέστησαν τα CRFs με δένδρα αποφάσεων (decision trees) επιτυγχάνοντας σημασιολογική κατάτμηση σε πραγματικό χρόνο.

4.2.1 Επισκόπηση Προτεινόμενης Προσέγγισης

Στην ενότητα 4.1 περιγράψαμε ορισμένες προσεγγίσεις στο πρόβλημα της κατάτμησης εικόνων και μάλιστα δύο (αλγόριθμοι κατάτμησης watershed και RSST) τις μελετήσαμε σε μεγαλύτερο βάθος προκειμένου να γίνει πιο κατανοητή η προτεινόμενη μεθοδολογία της σημασιολογικής κατάτμησης. Βασικός μας στόχος είναι να εμπλουτίσουμε τις τεχνικές κατάτμησης επέκτασης περιοχών (region growing) με σημασιολογικά χαρακτηριστικά, προκειμένου να επιλύσουμε προβλήματα υπερκατάτμησης ή υποκατάτμησης που συχνά παρατηρούνται και οφείλονται στην αδυναμία περιγραφής σημασιολογικών οντοτήτων με οπτικά μόνο χαρακτηριστικά. Πιο συγκεκριμένα, ο κύριος στόχος της σημασιολογικής επέκτασης περιοχών (semantic region growing - SRG) είναι να επιτύχει τόσο καλύτερη κατάτμηση της εικόνας, όσο και να χαρακτηρί-

σει και να κατατάξει τις επιμέρους περιοχές που συνθέτουν την εικόνα. Η καινοτομία του προτεινόμενου αλγορίθμου έγκειται στην τροποποίηση καθιερωμένων αλγορίθμων κατάτμησης ώστε να ενσωματώσουν χαρακτηριστικά ενδιάμεσου επιπέδου, όπως αυτά που περιγράφηκαν στην ενότητα 3.4.4.

Η ανάλυση με βάση την γνώση που παρουσιάσαμε στο κεφάλαιο 3 προηγείται της σημασιολογικής κατάτμησης. Συνοπτικά, η προσέγγιση έχει ως εξής: Η αρχική εικόνα κατατέμνεται με βάση έναν απλό ιεραρχικό αλγόριθμο ομοιογένειας ως προς το χρώμα και δημιουργεί πολλές μικρές περιοχές. Από αυτές τις περιοχές εξάγονται οι τοπικοί οπτικοί περιγραφείς οι οποίοι με κάποια μέθοδο μηχανικής μάθησης (ταίριασμα πρωτοτύπων, νευρωνικό δίκτυο, κ.ά.) αντιστοιχούνται με έννοιες αποθηκευμένες σε μία βάση γνώσης. Κατ' αυτόν τον τρόπο, κάθε περιοχή συσχετίζεται με ένα σύνολο πιθανών εννοιών. Η τοπολογία των περιοχών, οι οπτικοί περιγραφείς και το σύνολο των πιθανών εννοιών, αποθηκεύονται σε ένα σχεσιακό γράφο χαρακτηριστικών *ARG* ο οποίος είναι στην ουσία η είσοδος για το στάδιο της σημασιολογικής κατάτμησης. Πλέον η μετρική ομοιότητας δύο περιοχών υπολογίζεται με ασαφείς τελεστές των ασαφών συνόλων των εννοιών που έχουν συνδεθεί με τις περιοχές αυτές και τα ασαφή σύνολα εννοιών χρησιμοποιούνται ως κριτήριο για τη συγχώνευση γειτονικών περιοχών. Η προσέγγιση αυτή μπορεί να εφαρμοσθεί σε μία ποικιλία ιεραρχικών και μη αλγορίθμων συγχώνευσης περιοχών [1, 82, 77, 2, 37], με μικρές τροποποιήσεις κάθε φορά όσον αφορά τη μετρική ομοιότητας και τις συνθήκες τερματισμού.

Είναι σημαντικό να υπογραμμίσουμε ότι αυτή η προσέγγιση είναι ανεξάρτητη του αλγορίθμου κατάτμησης που θα επιλεγεί, αρκεί να ανήκει στη γενικότερη κατηγορία των αλγορίθμων επέκτασης περιοχών. Για να το καταστήσουμε σαφές εξετάσαμε δύο παραδοσιακές τεχνικές κατάτμησης, τον αλγόριθμο watershed [77] και τον αλγόριθμο recursive shortest spanning tree (RSST) [82]. Τροποποιήσαμε τις παραπάνω τεχνικές ώστε να ενεργούν στα ασαφή σύνολα που είναι αποθηκευμένα στον γράφο *ARG* με τρόπο παρόμοιο σαν να ενεργούσαν σε χαμηλού επιπέδου χαρακτηριστικά (όπως είναι το χρώμα, η υφή, κλπ.), όπως συμβαίνει παραδοσιακά. Οι δύο παραλλαγές ακολουθούν τις αντίστοιχες θεμελιώδεις αρχές των ορισμών των παραδοσιακών αλγορίθμων τους, ωστόσο έγιναν ορισμένες μετατροπές που θεωρήθηκαν απαραίτητες και οι οποίες αναφέρονται ρητά στο κείμενο που ακολουθεί.

4.2.2 Αλγόριθμος Σημασιολογικού Πλημμυρισμού

Στη δική μας σημασιολογική εκδοχή του μετασχηματισμού watershed (*SW*) έχουμε χρησιμοποιήσει μία καινοτόμο μέθοδο για την επιλογή των περιοχών-σπόρων, καθώς επίσης έχουμε τροποποιήσει το κριτήριο βάσει του οποίου δημιουργούνται οι υδατογραμμές. Τυπικά πρόκειται για ένα κριτήριο ομοιογένειας, το οποίο εξαρτάται άμεσα από το μετασχηματισμό της εικόνας πάνω στον οποίο εφαρμόζουμε τον αλγόριθμο watershed. Στην πιο συνήθη περίπτωση πρόκειται για τη τιμή της κλίσης της εικόνας (ή της γκρι συνιστώσας για έγχρωμες εικόνες). Όπως αναφέρθηκε και στην ενότητα 4.1.3, ένα από τα βασικά μειονεκτήματα του μετασχηματισμού watershed για κατάτμηση εικόνων είναι η υπερκατάτμηση και η δυσκολία επιλογής σημαδευτών ώστε να περιοριστεί ο αριθμός των περιοχών.

Στην προτεινόμενη τεχνική σημασιολογικού πλημμυρισμού *SW*, οι περιοχές που επιλέγονται ως σημαδευτές προκύπτουν από σημασιολογικά κριτήρια, και πιο συγκεκριμένα από το βαθμό εμπιστοσύνης που έχουμε ότι η περιοχή περιγράφεται από μία έννοια. Ο σχεσιακός γράφος χαρακτηριστικών *ARG*, που αποτελεί τη μορφή ανα-

παράστασης της εικόνας αντιστοιχίζει κάθε περιοχή s_a σε έναν κόμβο $v_a \in V$. Ο κόμβος v_a περιγράφεται από ένα διάνυσμα χαρακτηριστικών $\alpha(v_a)$ που περιέχει και το ασαφές σύνολο εννοιών \mathcal{L}_a , όπως αυτό προκύπτει από τη διαδικασία αρχικοποίησης του ARG.

Έστω $S \subseteq V$ το σύνολο των κόμβων των περιοχών που επιλέγονται για να χρησιμοποιηθούν ως σημαδευτές/σπόροι κατά την αρχικοποίηση του SW αλγορίθμου. Μία περιοχή s_a για να ανήκει στο σύνολο S (δηλαδή να ισχύει $v_a \in S$) οφείλει να πληροί τα παρακάτω δύο κριτήρια:

1. Το ύψος του ασαφούς συνόλου \mathcal{L}_a (ορισμός 2.3) να είναι μεγαλύτερο από ένα κατώφλι:

$$h(\mathcal{L}_a) > T_{seed} \quad (4.2)$$

2. Η περιοχή s_a να περιγράφεται από μία μόνο έννοια (κύρια έννοια):

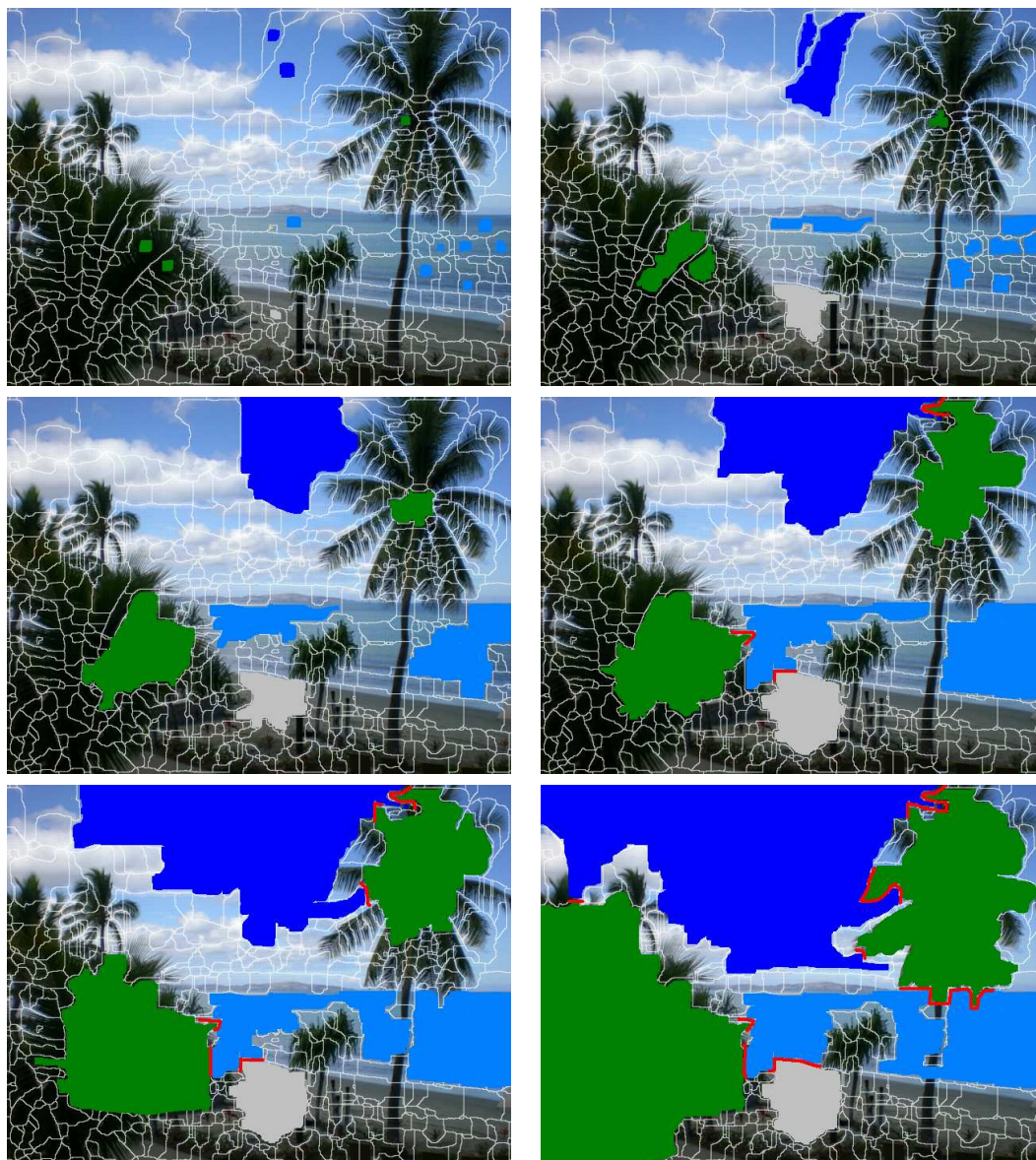
$$h(\mathcal{L}_a) > \sum_{c \in \{C-c^*\}} \mu_a(c), \quad \text{όπου } c^*: \mu_a(c^*) = h(\mathcal{L}_a) \quad (4.3)$$

Το πρώτο κριτήριο εξασφαλίζει ότι η περιοχή περιγράφεται, με αρκετά μεγάλο βαθμό εμπιστοσύνης, από μία τουλάχιστον έννοια. Η τιμή του κατώφλιου προσαρμόζεται αυτόματα σε κάθε εικόνα, με βάση το ιστόγραμμα των βαθμών εμπιστοσύνης όλων των περιοχών που την απαρτίζουν και για όλες τις έννοιες. Είναι επομένως λογικό να θεωρήσουμε ότι αυτές οι περιοχές έχουν ταξινομηθεί ορθά και μπορούν να αποτελέσουν τους σημασιολογικούς σημαδευτές για τον πλημμυρισμό των γειτονικών περιοχών. Ο δεύτερος περιορισμός εξασφαλίζει ότι η περιοχή περιγράφεται πλήρως από μία μόνο έννοια, απορρίπτοντας περιπτώσεις που περισσότερες έννοιες έχουν συσχετιστεί (συνήθως εξαιτίας λανθασμένης ταξινόμησης) με μία περιοχή. Τα δύο αυτά κριτήρια χρησιμοποιούνται για την ορθή επιλογή των περιοχών ως σημαδευτές για τη συγκεκριμένη έννοια c^* .

Μία επαναλαμβανόμενη διαδικασία ελέγχει κάθε περιοχή-σημαδευτή $s \in S$, και εντοπίζει τις άμεσα χωρικά γειτονικές περιοχές, με βάση την 8-γειτονιά. Έστω $r \in N_s$ μία γειτονική περιοχή του s , και $p(r)$ η αναδρομική συνάρτηση διάδοσης, τέτοια ώστε $s = p(r)$. Συγκρίνουμε τα ασαφή σύνολα των περιοχών r και $p(r)$, \mathcal{L}_r και $\mathcal{L}_{p(r)}$, αντίστοιχα, στοιχείο προς στοιχείο και για κάθε κοινή έννοια μετράμε το βαθμό συμμετοχής της συγκεκριμένης έννοιας στην περιοχή r , $\mu_r(c)$. Εάν ο βαθμός αυτός είναι μεγαλύτερος ενός κατώφλιου συγχώνευσης, τότε θεωρούμε ότι η περιοχή r είναι σημασιολογικά παρόμοια με το διασπορέα της $p(r)$. Το κατώφλι συγχώνευσης ορίζεται σε σχέση με την απόσταση της περιοχής r από τον αρχικό σημαδευτή s : $\mu_r(c) > K^n T_{merge}$, όπου η παράμετρος K είναι μία σταθερά λίγο μεγαλύτερη της μονάδας, έτσι ώστε το κατώφλι σε κάθε επανάληψη n να μεγαλώνει μη-γραμμικά σε σχέση με την απόσταση από το σημαδευτή. Η εν λόγω απόσταση δε μετρείται σε εικονοστοιχεία αλλά στον αριθμό των βημάτων-επανάληψεων του αλγορίθμου για να προσπελαστεί η εν λόγω περιοχή.

Στη συνέχεια η περιοχή r προστίθεται σε ένα νέο σύνολο περιοχών M_s^n , όπου το n συμβολίζει το βήμα της επαναληπτικής διαδικασίας, και $M_s^0 \triangleq S$, $M_s^1 \triangleq N_s$, κλπ. Στην επόμενη επανάληψη του αλγορίθμου η επιλογή των νέων σημαδευτών θα γίνει από το σύνολο M_s^n . Μετά τη συγχώνευση ο αλγόριθμος επαναυπολογίζει τους βαθμούς συμμετοχής όλων των εννοιών για τη νέα περιοχή \hat{r} του συνόλου $\mathcal{L}_{\hat{r}}$ ως την ασαφή τομή των συνόλων \mathcal{L}_r και $\mathcal{L}_{p(r)}$:

$$\mathcal{L}_{\hat{r}} = t(\mathcal{L}_{p(r)}, \mathcal{L}_r) = \mathcal{L}_{p(r)} \wedge \mathcal{L}_r \quad (4.4)$$



Σχήμα 4.6: Σχηματική αναπαράσταση του σημασιολογικού πλημμυρισμού εικόνας.

Η παραπάνω διαδικασία επαναλαμβάνεται έως το βήμα n που ικανοποιείται το κριτήριο τερματισμού: τα ασαφή σύνολα για όλες τις περιοχές-σημαδευτές να είναι κενά, δηλαδή $M_s^n = \emptyset$. Σε αυτό το σημείο πρέπει να υπογραμμίσουμε πως όταν εξετάζονται οι γείτονες μίας περιοχής, οι περιοχές που έχουν εξεταστεί σε προηγούμενο βήμα του αλγορίθμου εξαιρούνται, ούτως ώστε κάθε περιοχή προσπελάζεται μία μόνο φορά, από την κοντινότερη περιοχή-σημαδευτή, όπως καθορίζει η γεωμετρία του γράφου ARG .

Στο σχήμα 4.6 παρουσιάζουμε ορισμένα βήματα του σημασιολογικού πλημμυρισμού. Καταρχήν η εικόνα έχει κατατμηθεί σε πάρα πολλές μικρές περιοχές με τον αλγόριθμο watershed. Από τις περιοχές αυτές σχηματίζουμε το σχεσιακό γράφο χαρακτηριστικών ARG και εντοπίζουμε τις περιοχές-σημαδευτές, οι οποίες πληρούν τις προαναφερθείσες προϋποθέσεις. Στο ίδιο σχήμα έχουμε μαρκάρει τις περιοχές αυτές με ξεχωριστό χρώμα ανάλογα με την κύρια έννοια c^* που τις αντιστοιχεί (σκούρο μπλε η έννοια ΟΥΡΑΝΟΣ, ανοιχτό μπλε η έννοια ΘΑΛΛΑΣΑ, πράσινο η έννοια ΒΛΑΣΤΗΣΗ και γκρι η έννοια ΑΜΜΟΣ). Έπειτα εξετάζονται οι γειτονικές περιοχές, δηλαδή το σύνολο $\{N_{s_a}\}$, $\forall s_a \in S$ και εφόσον ικανοποιείται το κριτήριο συγχώνευσης δύο

γειτονικών περιοχών, τότε αυτές συνενώνονται και ενημερώνονται καταλλήλως τα ασαφή σύνολα εννοιών των περιοχών αυτών. Η διαδικασία αυτή είναι ίδια για όλες τις περιοχές-σημαδευτές και συνεχίζεται έως ότου συναντηθούν δύο περιοχές που προέρχονται από διαφορετικό σημαδευτή (κόκκινη διαχωριστική γραμμή στο σχήμα) ή δεν καλύπτεται το κατώφλι συγχώνευσης $K^n \cdot T_{merge}$.

Με το πέρας της εκτέλεσης του αλγορίθμου *SW*, ορισμένες περιοχές θα έχουν συγχωνευτεί με τους σημασιολογικά κοντινότερους σημαδευτές, ενώ άλλες θα έχουν μείνει ανεπηρέαστες. Προκειμένου να μελετήσουμε αυτές τις περιοχές, εκτελούμε τον αλγόριθμο *SW* στο νέο *ARG*, δημιουργώντας νέες περιοχές-σημαδευτές. Αυτή η ιεραρχική στρατηγική δεν χρειάζεται επιπλέον παραμετροποίηση σε σχέση με τον αλγόριθμο *SW*, καθώς οι νέοι σημαδευτές δημιουργούνται αυτόματα (με μικρότερη τιμή για το κατώφλι T_{seed}). Όπως είναι αναμενόμενο, οι περιοχές που δημιουργούνται κατά την πρώτη εκτέλεση του *SW* έχουν μεγαλύτερη βεβαιότητα για τα όρια τους και για την ορθή σημασιολογική τους αναπαράσταση, σε σύγκριση με τις περιοχές που δημιουργούνται στις επόμενες εκτελέσεις του *SW*. Παρόλα αυτά δεν είναι μειονέκτημα για τον αλγόριθμο, καθώς οι περιοχές που δημιουργούνται κατά το δεύτερο και τρίτο βήμα είναι καλύτερες από αυτές που είχαν προκύψει με τον αρχικό αλγόριθμο κατάτμησης και επιπρόσθετα διατηρείται η ασάφεια στην ταξινόμηση των περιοχών στο τελικό αποτέλεσμα, που είναι χρήσιμη σε επόμενα στάδια επεξεργασίας.

4.2.3 Αλγόριθμος Σημασιολογικού RSST

Ο παραδοσιακός αλγόριθμος κατάτμησης RSST [82] ανήκει στην κατηγορία των κάτωθεν (bottom-up) αλγορίθμων. Γενικά, αυτού του είδους οι αλγόριθμοι ξεκινάνε ενώνοντας επαναληπτικά γειτονικά εικονοστοιχεία και περιοχές με βάση ένα ή παραπάνω κριτήρια ομοιότητας, έως ότου ικανοποιηθεί η συνθήκη τερματισμού τους και οι συνεκτικές περιοχές που δημιουργούνται αποτελούν την τελική κατάτμηση της εικόνας. Πιο συγκεκριμένα, και όπως περιγράψαμε στην ενότητα 4.1.2, ο αλγόριθμος RSST έχει εσωτερικά αποθηκευμένα, είτε σε μορφή λίστας είτε σε δομή γράφου, όλα τα εικονοστοιχεία που απαρτίζουν την εικόνα μαζί με πληροφορίες όπως η φωτεινότητα, το χρώμα τους, η χρωματική αντίθεση από τα γειτονικά τους κ.ά.. Με βάση το κριτήριο ομοιότητας που έχει οριστεί κάθε φορά, υπολογίζεται η απόσταση για κάθε ζεύγος γειτονικών εικονοστοιχείων. Η γειτονιά μπορεί να οριστεί με οποιονδήποτε αυθαίρετο τρόπο, καλό είναι πάντως να είναι μικρή για να μην επιβαρύνει την ταχύτητα του αλγορίθμου.

Η χρήση ενός γράφου *ARG*, όπως αυτός περιγράφηκε στην ενότητα 2.3.1, είναι κατάλληλη ως δομή για την εκτέλεση του RSST αλγορίθμου. Κατά την αρχικοποίηση του RSST κάθε εικονοστοιχείο αντιστοιχίζεται σε έναν κόμβο του *ARG* και κάθε ζεύγος γειτονικών εικονοστοιχείων σε μία ακμή. Οι ακμές ταξινομούνται ως προς το βάρος τους, που προκύπτει με βάση το κριτήριο ομοιότητας (για παράδειγμα μπορεί να χρησιμοποιηθεί η χρωματική ανομοιότητα των συνδεόμενων κόμβων βάσει της αντίστροφης κανονικοποιημένης ευκλείδειας απόστασης των χρωματικών συνιστωσών των). Επιλέγεται η ακμή με το ελάχιστο βάρος για να αφαιρεθεί από τον *ARG* και να συγχωνευτούν οι παρακείμενοι κόμβοι σε έναν. Ύστερα από κάθε βήμα τα χαρακτηριστικά της προκύπτουσας περιοχής (όπως π.χ. το μέσο χρώμα) επαναυπολογίζονται. Επιπλέον, ο RSST επαναυπολογίζει και τα βάρη των ακμών που επηρεάστηκαν και επαναταξινομεί τις ακμές του *ARG*, έτσι ώστε σε κάθε βήμα να επιλέγεται με ορθό τρόπο η ακμή με το μικρότερο βάρος. Η διαδικασία αυτή συνεχίζει έως ότου ικανο-

ποιηθεί κάποια από τις συνθήκες τερματισμού, οι οποίες συνήθως είναι μία μέγιστη τιμή για το βάρος μίας ακμής που μπορεί να διαγραφεί είτε ο αριθμός των κόμβων στον τελικό γράφο.

Ακολουθώντας τις συμβάσεις και τα σύμβολα που έχουμε χρησιμοποιήσει έως τώρα, παρουσιάζουμε μία τροποποιημένη έκδοση του RSST αλγορίθμου, την οποία ονομάζουμε σημασιολογικό RSST (*SRSST*). Η βασική διαφοροποίηση του *SRSST* από τον παραδοσιακό αλγόριθμο έγκειται στον υπολογισμό του κριτηρίου ομοιότητας δύο περιοχών. Επίσης, σε αντιδιαστολή με τον αλγόριθμο *SW* που περιγράφηκε στην προηγούμενη παράγραφο, ο *SRSST* δεν χρησιμοποιεί περιοχές-σημαδευτές κατά την αρχικοποίηση του. Το βάρος μίας ακμής e_{ab} του *ARG* υπολογίζεται από τα ασαφή σύνολα των παρακείμενων κόμβων a και b :

$$w(e_{ab}) = 1 - h(\mathcal{L}_a \wedge \mathcal{L}_b) \quad (4.5)$$

όπου \wedge είναι η ασαφής τομή των ασαφών συνόλων εννοιών $\mathcal{L}_a, \mathcal{L}_b$ που περιγράφουν τις περιοχές a, b αντιστοίχως και $h(\cdot)$ είναι ο τελεστής ύψους του ασαφούς συνόλου. Καθώς το ύψος ενός ασαφούς συνόλου ανήκει στο διάστημα $[0, 1]$ τότε αρκεί να το αφαιρέσουμε από τη μονάδα για να έχουμε το κανονικοποιημένο μέτρο ανομοιότητας που αναπαριστάται με το βάρος της ακμής. Καθώς ο τελεστής \wedge , όπως όλοι οι ασαφείς τελεστές τομής, ικανοποιεί τις ιδιότητες της αντιμεταθετικότητας και της προσεταιριστικότητας είναι εξασφαλισμένο ότι η σειρά των ορισμάτων στην εξίσωση (4.5) είναι αδιάφορη και έτσι ταξινομούνται όλες οι ακμές του *ARG* χωρίς να παίζει ρόλο η σειρά.

Ας εξετάσουμε λεπτομερώς ένα βήμα του *SRSST* αλγορίθμου. Καταρχήν, εντοπίζεται η ακμή με το ελάχιστο βάρος στον *ARG*: $e_{ab}^* = \arg \min_{e_{ab} \in E} (w(e_{ab}))$. Οι κόμβοι του γράφου v_a και v_b που αντιστοιχούν στις περιοχές a και b ενώνονται και σχηματίζουν μία νέα περιοχή, έστω \hat{a} . Ο *ARG* ανανεώνεται κατάλληλα ώστε να απεικονίζει τη νέα κατάτμηση της εικόνας. Η διαδικασία της ανανέωσης του *ARG* περιλαμβάνει τις ακόλουθες ενέργειες:

1. Υπολογισμός του ασαφούς συνόλου $\mathcal{L}_{\hat{a}}$ επανεκτιμώντας όλους τους βαθμούς εμπιστοσύνης για όλες τις έννοιες:

$$\mu_{\hat{a}}(c) = \frac{|a| \cdot \mu_a(c) + |b| \cdot \mu_b(c)}{|a| + |b|}, \quad \forall c \in \mathcal{C} \quad (4.6)$$

Η ποσότητα $|\cdot|$ είναι το μέγεθος της περιοχής, δηλαδή ο αριθμός των εικονοστοιχείων που απαρτίζουν την περιοχή.

2. Προσαρμογή των ακμών του *ARG*:

- (a) Αφαίρεση της ακμής e_{ab} .
- (b) Επανεκτίμηση των βαρών των ακμών e που επηρεάστηκαν από τη συγχώνευση. Αυτές είναι το σύνολο των ακμών που ένας από τους δύο προσκείμενους κόμβους έλαβε μέρος στη συγχώνευση των περιοχών a και b : $e \in E_a \cup E_b$.

Τα παραπάνω βήματα επαναλαμβάνονται μέχρις ότου το βάρος της ακμής e^* να είναι μεγαλύτερο από ένα κατώφλι: $w(e^*) > T_w$. Το T_w υπολογίζεται στην αρχή του αλγορίθμου βάσει του αθροιστικού ιστογράμματος των βαρών όλων των ακμών E του *ARG* (παρόμοια με τον υπολογισμό του T_{seed}).

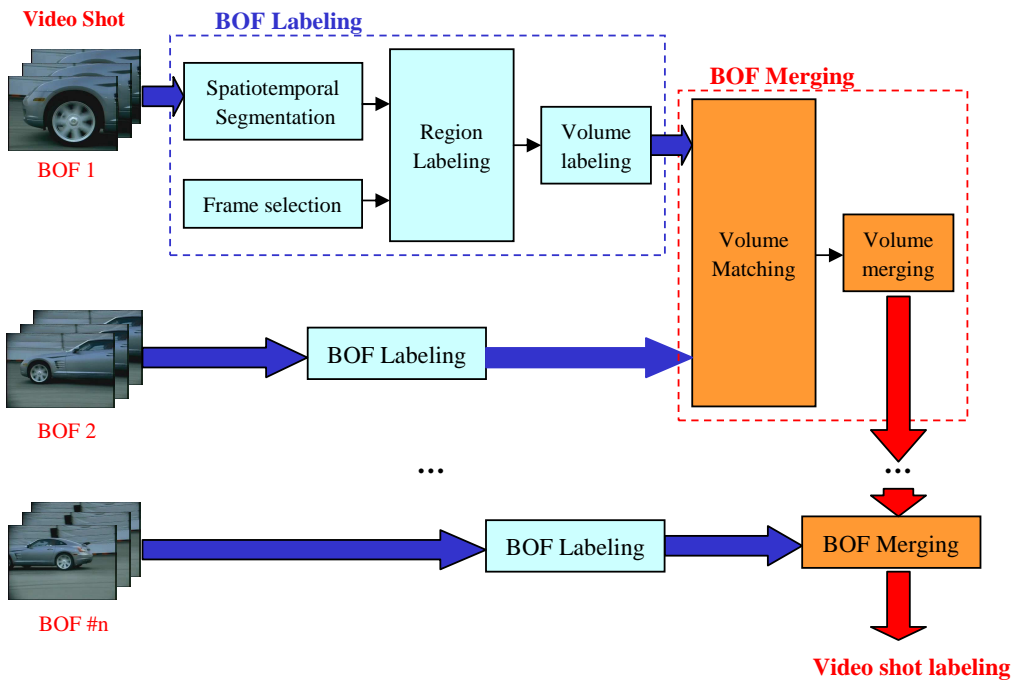
4.3 Σημασιολογική Κατάτμηση Βίντεο

Τα τελευταία χρόνια η παραγωγή πολυμεσικού περιεχομένου έχει απλοποιηθεί πάρα πολύ διευκολύνοντας τους απλούς χρήστες να δημιουργούν προσωπικές συλλογές βίντεο και να τις μοιράζονται. Ενδεικτικά, σε πρόσφατη έρευνα αναφέρθηκε ότι στο δικτυακό χώρο youtube.com συλλέγονται 20 ώρες βίντεο κάθε λεπτό. Αυτός ο τεράστιος όγκος δεδομένων, μαζί με την ανάγκη για αποδοτική συγκρότηση, δεικτοδότηση και αναζήτηση που έχει δημιουργηθεί, καθιστούν απαραίτητη την ανάπτυξη αλγορίθμων που να αναλύουν σημασιολογικά το περιεχόμενο των βίντεο.

Σε συνέχεια της προηγούμενης ενότητας, επεκτείνουμε τη σημασιολογική κατάτμηση σε ακολουθίες βίντεο, προτείνοντας έναν υβριδικό αλγόριθμο χωροχρονικής κατάτμησης και σημασιολογικού χαρακτηρισμού πλάνων βίντεο. Ένα πλάνο (video shot) ορίζεται ως η δομική μονάδα του βίντεο που αποτελείται από συνεχόμενα καρέ που προκύπτουν από μία συγκεκριμένη κάμερα με εγγραφή χωρίς διακοπή [19]. Η χωροχρονική κατάτμηση τεμαχίζει ένα πλάνο σε συνεχόμενους 3Δ όγκους (2Δ+χρόνος) οι οποίοι είναι ομοιογενείς ως προς ένα σύνολο χαρακτηριστικών. Αυτοί οι όγκοι αποτελούν το μέσο, κατ' αναλογία με τις περιοχές στην περίπτωση των εικόνων, για τη σημασιολογική ταξινόμηση και διάδοση των εννοιών σε ένα πλάνο.

Έχουν προταθεί αρκετές τεχνικές για τη χωροχρονική κατάτμηση βίντεο σε όγκους. Οι 3Δ προσεγγίσεις έχουν ως είσοδο ένα πλάνο (ή ισοδύναμα το σύνολο των καρέ (frames) που το αποτελούν) και βελτιστοποιώντας κάποιο κριτήριο εξάγουν συνεκτικούς όγκους [63]. Πιο συγκεκριμένα, οι όγκοι έχουν μοντελοποιηθεί με μείγμα μοντέλων gaussian κατανομών (gaussian mixtures models) για το χρώμα και τη θέση [47] ή θεωρήθηκαν ως μικρά μπαλώματα (patches) που κινούνται γραμμικά [32]. Επίσης τεχνικές οπτικής προσοχής έχουν χρησιμοποιηθεί για την εξαγωγή των σημαντικών όγκων, εντοπίζοντας πρώτα 3Δ σημεία ενδιαφέροντος [101]. Το βασικότερο πρόβλημα των 3Δ προσεγγίσεων στη χωροχρονική κατάτμηση είναι η πολυπλοκότητά τους και το αυξημένο υπολογιστικό τους κόστος. Προκειμένου να μειωθεί η πολυπλοκότητα, έχουν προταθεί προσεγγίσεις που χρησιμοποιούν την αναπαράσταση 2Δ+χρόνος και είναι πιο αποδοτικές, εις βάρος όμως της ακρίβειας στην εξαγωγή των όγκων [41].

Η ενσωμάτωση σημασιολογικής πληροφορίας στη διαδικασία της χωροχρονικής κατάτμησης θέτει δύο κύριες προκλήσεις. Πρώτον, η αναγνώριση εννοιών για περιοχές, η οποία επιτυγχάνεται με την εξαγωγή και το ταίριασμα οπτικών χαρακτηριστικών, επιφέρει ένα πρόσθετο υπολογιστικό κόστος, το οποίο είναι ακόμα μεγαλύτερο όταν πρόκειται για 3Δ χαρακτηριστικά. Δεύτερον, η αξιοπιστία των σημασιολογικών χαρακτηριστικών εξαρτάται από πάρα πολλές παραμέτρους και δεν είναι πάντοτε αρκετά μεγάλη για να είναι χρήσιμη. Οι δύο αυτοί παράγοντες υποδεικνύουν ότι η χρήση σημασιολογικών χαρακτηριστικών στα πρώτα στάδια ενός αλγορίθμου κατάτμησης είναι μη αποδοτική και αναποτελεσματική, αν όχι και παραπλανητική. Επομένως, η προσέγγιση που περιγράφεται στις επόμενες ενότητες, προκειμένου να ανταποκριθεί στις δύο βασικές προαναφερθείσες προκλήσεις, εισάγει δύο καινοτόμες ιδέες: πρώτον, η ενσωμάτωση της σημασιολογικής πληροφορίας στο στάδιο της κατάτμησης γίνεται όταν υπάρχει σχετικά μικρός αριθμός όγκων και δεύτερον, ότι η εξαγωγή των όγκων δεν γίνεται από όλο το πλάνο, αλλά προηγείται η χρονική κατάτμηση του σε μπλοκ από καρέ (block of frames - BoF) και στο τέλος προτείνεται μεθοδολογία για τη σημασιολογική συγχώνευση των BoFs ενός πλάνου.



Σχήμα 4.7: Διάγραμμα λειτουργίας σημασιολογικής κατάτμησης βίντεο.

4.3.1 Επισκόπηση Μεθοδολογίας

Η προτεινόμενη μεθοδολογία σημασιολογικής κατάτμησης βίντεο απεικονίζεται στο σχήμα 4.7. Οι ακολουθίες βίντεο θεωρούμε ότι αποτελούνται από ένα πλάνο, διαφορετικά πρέπει να προηγηθεί ένας αλγόριθμος διαχωρισμού πλάνων (σύμφωνα με τη σχετική βιβλιογραφία, αξιόπιστοι αλγόριθμοι επιτυγχάνουν ακρίβεια άνω του 90% [19]). Εξαιτίας επικαλύψεων, σκίασης, αλλαγή οπτικής γωνίας, κίνησης της κάμερας κ.ά. τα αντικείμενα έχουν μεγάλη χωρική και χρονική απόκλιση, δυσχεραίνοντας τον εντοπισμό και την παρακολούθησή τους. Η αποδόμηση ενός πλάνου σε μπλοκ από λίγα καρέ (BoF) εκτός από τη μείωση της υπολογιστικής πολυπλοκότητας, μειώνει και αυτή τη χωροχρονική απόκλιση.

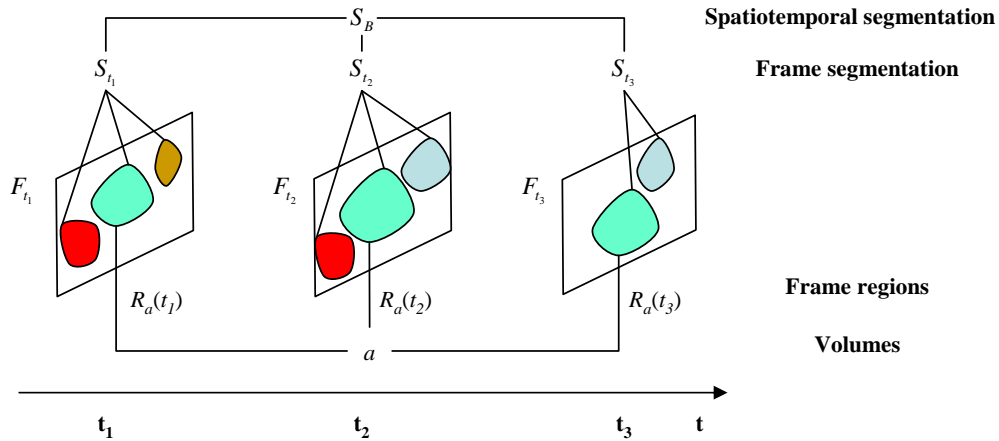
Η σημασιολογική κατάτμηση βίντεο αποτελείται από δύο βασικά στάδια: πρώτον, από τη σειριακή επεξεργασία για όλα τα BoFs που συνίσταται στην επισήμανση (labeling) των όγκων με έννοιες και τη συγχώνευση αυτών με οπτικά και σημασιολογικά κριτήρια (intra-BoF ανάλυση) και δεύτερον, από το ταίριασμα και τη συγχώνευση όγκων μεταξύ διαδοχικών BoFs (inter-BoF ανάλυση). Το πρώτο στάδιο απεικονίζεται μέσα στο πλαίσιο με τη μπλε διακεκομμένη γραμμή στο σχήμα 4.7. Πρώτη διαδικασία μέσα σε αυτό το πλαίσιο είναι η χωροχρονική κατάτμηση για την αποδόμηση του βίντεο σε ένα σύνολο συνεκτικών όγκων. Το αποτέλεσμα είναι ένας 2Δ +χρόνος χάρτης κατάτμησης, ο οποίος δειγματοληπτείται χρονικά για να παραχθούν οι μάσκες κατάτμησης ανά καρέ, κάθε μία εκ των οποίων αποτελείται από ένα σύνολο μη επικαλυπτόμενων περιοχών. Οι περιοχές αυτές επισημαίνονται με έννοιες (region labeling), με ακριβώς την ίδια διαδικασία που περιγράφηκε στο κεφάλαιο 3. Οι έννοιες που αποδόθηκαν στις περιοχές, αντιστοιχίζονται στους όγκους (volume labeling), από τους οποίους έχουν προκύψει, εξ αρχής, οι περιοχές. Τέλος, ακολουθεί η διαδικασία σημασιολογικής επέκτασης όγκων, με παρόμοιο τρόπο όπως αναλύθηκε νωρίτερα στην ενότητα 4.2, με κυριότερη διαφορά ότι ο σχεσιακός γράφος

χαρακτηριστικών ARG έχει και χρονικές ακμές.

Το δεύτερο στάδιο της ανάλυσης (απεικονιζόμενο με το πλαίσιο με την κόκκινη διακεκομμένη γραμμή στο σχήμα 4.7), έχει ως στόχο να συνδέσει τους όγκους μεταξύ διαδοχικών BoFs και να παράγει συνεχείς, συνεκτικούς όγκους εντός του πλάνου, που να αντιστοιχούν στη χωροχρονική τροχιά του κάθε εντοπισμένου και αναγνωρισμένου αντικειμένου. Το ταίριασμα των όγκων (volume matching) γίνεται με βάση τη σημασιολογική τους ερμηνεία και αποτελεί το κριτήριο για τη συγχώνευσή δύο διαδοχικών BoFs (BoFs merging). Η διαδικασία επαναλαμβάνεται για όλα τα ζεύγη διαδοχικών BoFs, ολοκληρώνοντας τη σημασιολογική χωροχρονική κατάτμηση του πλάνου.

4.3.2 Αναπαράσταση Βίντεο με Γράφο

Ακολουθώντας το πρότυπο MPEG-7 για την περιγραφή πολυμέσων, ένα πλάνο βίντεο αποδομείται σε τμήματα βίντεο με ιεραρχικό τρόπο. Καταρχήν, ένα πλάνο χωρίζεται σε M μπλοκ από καρέ (BoFs) B_i ($i \in [1, M]$), κάθε ένα εκ των οποίων αποτελείται από διαδοχικά καρέ F_t , $t \in [1, |B_i|]$. Η χωροχρονική κατάτμηση τεμαχίζει κάθε B_i σε ένα σύνολο 3D περιοχών του βίντεο (όγκων) S_{B_i} . Κάθε όγκος $a \in S_{B_i}$ υποδιαιρείται χρονικά σε 2D περιοχές $R_a(t)$, $F_t \in B_i$. Τέλος, η κατάτμηση του καρέ τη χρονική στιγμή t ορίζεται ως η ένωση των περιοχών όλων των όγκων που τέμνουν το καρέ F_t : $S_t = \bigcup_{a \cap F_t \neq \emptyset} R_a(t)$. Τα στοιχεία που συνθέτουν ένα BoF απεικονίζονται στο σχήμα 4.8.



Σχήμα 4.8: Χωρική και χρονική αποδόμηση ενός BoF B_i .

Η αναπαράσταση ενός τμήματος βίντεο, όπως είναι ένα BoF, μπορεί να γίνει με χρήση ενός σχεσιακού γράφου χαρακτηριστικών ARG , όπως αυτός ορίστηκε στην ενότητα 2.3.1. Αν ορίσουμε S_{B_i} την κατάτμηση ενός BoF B_i , τότε ένας όγκος $a \in S_{B_i}$ αναπαριστάται στον γράφο $G \equiv (V, E)$ με τον κόμβο $v_a \in V$, όπου $v_a \equiv \langle a, \mathcal{D}_a, \mathcal{L}_a \rangle$. Η ποσότητα \mathcal{D}_a είναι το σύνολο των οπτικών περιγραφέων για τον όγκο a , ενώ η έκφραση \mathcal{L}_a είναι το ασαφές σύνολο εννοιών που περιγράφουν σημασιολογικά τον όγκο a .

Δύο (χωροχρονικά) γειτονικοί όγκοι $a, b \in S_{B_i}$ συνδέονται με μία ακμή του γράφου $e_{ab} \in E$ με διάνυσμα χαρακτηριστικών $w(e_{ab}) = \langle s_{ab}^{\mathcal{D}}, s_{ab}^{\mathcal{L}} \rangle$. Η ποσότητα $s_{ab}^{\mathcal{D}}$ είναι η οπτική ομοιότητα των δύο όγκων a και b όπως υπολογίζεται από το διάνυσμα των οπτικών περιγραφέων \mathcal{D}_a και \mathcal{D}_b . Επειδή υπάρχουν διαφορετικές μετρικές απόστασης για κάθε έναν οπτικό περιγραφέα (ειδικά για τους περιγραφείς εκτός του προτύπου

MPEG-7), επιλέγουμε τον κανονικοποιημένο γραμμικό συνδυασμό αυτών των αποστάσεων και υπολογίζουμε έτσι τη συνολική οπτική απόσταση των δύο υπό εξέταση όγκων. Χρησιμοποιούμε συναρτήσεις κανονικοποίησης απόστασης για κάθε περιγραφέα $\{n_d\}$, $d \in \mathcal{D}$ και συντελεστές βάρους α_d ανάλογα με τη σπουδαιότητα του κάθε περιγραφέα:

$$s_{ab}^{\mathcal{D}} = \sum_{d \in \mathcal{D}} \alpha_d n_d \quad (4.7)$$

Η ποσότητα $s_{ab}^{\mathcal{L}}$ αντικατοπτρίζει τη σημασιολογική ομοιότητα των δύο όγκων a και b και υπολογίζεται από τα δύο ασαφή σύνολα εννοιών \mathcal{L}_a και \mathcal{L}_b με την ακόλουθη εξίσωση:

$$s_{ab}^{\mathcal{L}} = \sup_{c_i \in \mathcal{C}} (t(\mathcal{L}_a, \mathcal{L}_b)), \quad a \in S, b \in N_a \quad (4.8)$$

N_a είναι το σύνολο των γειτονικών όγκων του a και t είναι μία τ-νόρμα δύο ασαφών συνόλων. Διαισθητικά, η εξίσωση (4.8) διατυπώνει ότι η σημασιολογική ομοιότητα $s_{ab}^{\mathcal{L}}$ είναι ο μέγιστος βαθμός, όπως συνεπάγεται από τη γνώση μας, ότι οι όγκοι a και b μοιράζονται την ίδια έννοια.

4.3.3 Ανάλυση ενός BoF

Ο πιο εύλογος τρόπος για να αντιστοιχίσουμε μία έννοια σε ένα τμήμα βίντεο θα ήταν να απομονώσουμε την 3Δ περιοχή του βίντεο (όγκος), να εξάγουμε τα οπτικά χωροχρονικά χαρακτηριστικά του και να χρησιμοποιήσουμε κάποια τεχνική μηχανικής μάθησης για να τα ταιριάξουμε με έννοιες από μία βάση γνώσης. Δυστυχώς η προσέγγιση αυτή αντιμετωπίζει σοβαρά προβλήματα εξαιτίας του μεγάλου υπολογιστικού κόστους που έχουν οι διαδικασίες εξαγωγής και ταιριάσματος χαρακτηριστικών από όγκους μέσα από μία ακολουθία βίντεο, ή η σημασιολογική ανάλυση όλων των καρέ ξεχωριστά.

Προτείνουμε λοιπόν μία εναλλακτική προσέγγιση για την επισήμανση (labeling) των αντικειμένων, εντός του χρονικού πλαισίου που καθορίζει η διάρκεια ενός μπλοκ. Επιλέγουμε ορισμένα μόνο καρέ για να χαρακτηρίσουμε σημασιολογικά τις περιοχές από τις οποίες αποτελούνται και αξιοποιούμε τη χωροχρονική κατάτμηση για τη διάδοση των εννοιών από τις 2Δ περιοχές μέσα στους 2Δ+χρόνο όγκους. Οι επόμενες ενότητες αναλύουν την επιλογή των χαρακτηριστικών καρέ, τη σημασιολογική περιγραφή των όγκων και την επέκταση και συγχώνευση των όγκων μέσα σε ένα BoF.

4.3.3.1 Επιλογή Καρέ

Η επιλογή ορισμένων καρέ από ένα μπλοκ καρέ (BoF) γίνεται για να ελαττωθεί το υπολογιστικό κόστος της εξαγωγής οπτικών περιγραφών και της ταξινόμησης των περιοχών, όπως αυτές προκύπτουν από την τομή της χωροχρονικής κατάτμησης με το συγκεκριμένο καρέ. Η επιλογή πολλών χαρακτηριστικών καρέ θα οδηγήσει σε καλύτερη περιγραφή του BoF, ενώ αντίθετα η επιλογή λίγων χαρακτηριστικών καρέ θα οδηγήσει σε πιο αποδοτική και γρήγορη επεξεργασία. Στην ακραία περίπτωση που επιλεγεί ένα μόνο καρέ, τότε σε πολλούς όγκους δε θα αποδοθούν έννοιες και δε θα επιτευχθεί η σημασιολογική συγχώνευσή τους.

Επιλέγουμε ένα σύνολο από καρέ, έστω T , και τις αντίστοιχες μάσκες κατάτμησης $S_T = \{S_t\}$, $t \in T$ και υπολογίζουμε το άνοιγμα των όγκων που περιλαμβάνονται στο

T . Με τον όρο άνοιγμα ενός όγκου a , εννοούμε τον αριθμό των καρέ στα οποία έχει εντοπιστεί ο a , ή διαφορετικά το σύνολο των καρέ με το οποίο έχει μη κενή τομή. Δεδομένου ενός σταθερού μεγέθους του συνόλου T , επιλέγουμε το σύνολο T^* το οποίο μεγιστοποιεί το συνολικό άνοιγμα των εντοπισμένων όγκων σε ένα BoF B_i :

$$T^* = \arg \max_{T \in \mathcal{F}^{|T|}(F_t)} \sum_{\forall a \in S_{B_i}, a \cap S_T \neq \emptyset} |a| \quad (4.9)$$

Η παραπάνω παράσταση $\mathcal{F}^{|T|}(F_t)$, είναι η οικογένεια των υποσυνόλων πληθυκότητας $|T|$ από το σύνολο F_t των καρέ του BoF B_i . Επίσης, S_{B_i} είναι το σύνολο των κατατμήσεων όλων των καρέ του B_i και $|a|$ είναι το μέγεθος του όγκου a . Το κριτήριο της σχέσης (4.9) υπερτερεί σε σχέση με την επιλογή συγκεκριμένων καρέ καθώς παρέχει την καλύτερη επιλογή όγκων με τη μέγιστη δυνατή κάλυψη σε όλο το BoF.

4.3.3.2 Υπολογισμός Σημασιολογικής και Συντακτικής Περιγραφής Όγκου

Στις προηγούμενες ενότητες δείξαμε πως γίνεται η επιλογή των καρέ T^* από ένα μπλοκ B_i και η εξαγωγή των περιοχών $R_a(t)$ (όπου $a \in S_{B_i}$ και t ο δείκτης του καρέ $F_t \in B_i$) μέσω της τομής του όγκου a με το καρέ $F_t \in B_i$. Χρησιμοποιώντας το μηχανισμό ταξινόμησης περιοχών, που περιγράψαμε στο κεφάλαιο 3, αντιστοιχούμε σε κάθε περιοχή $R_a(t)$ ένα ασαφές σύνολο εννοιών $\mathcal{L}_{R_a(t)}$. Το ασαφές σύνολο εννοιών \mathcal{L}_a του όγκου a , συνυπολογίζεται από τα ασαφή σύνολα των περιοχών $R_a(t)$ για όλα τα καρέ $F_t \in B_i$ που ανήκουν στον όγκο a με τον παρακάτω τελεστή συναθροίσις (aggregation operator):

$$\mathcal{L}_a = \frac{\sum_{t \in T^*} w(R_a(t)) \mathcal{L}_{R_a(t)}}{\sum_{t \in T^*} w(R_a(t))} \quad (4.10)$$

Ο εν λόγω τελεστής συναθροίζει τους βαθμούς συμμετοχής κάθε έννοιας από τις περιοχές $R_a(t)$ που ανήκουν στον όγκο a με διαφορετικό βάρος $w(R_a(t))$ για κάθε περιοχή, το οποίο βάρος καθορίζεται από τη χρονική συνοχή και συνέχεια των περιοχών.

Εκτός της σημασιολογικής επισήμανσης, οι όγκοι χαρακτηρίζονται και από χαμηλού επιπέδου οπτικούς περιγραφείς. Αν και οι περισσότεροι περιγραφείς του προτύπου MPEG-7 δημιουργήθηκαν για την περιγραφή 2Δ περιοχών, ορισμένοι μπορούν να επεκταθούν εύκολα και σε 3Δ όγκους χωρίς την ανάγκη επαναυπολογισμού τους, αλλά χρησιμοποιώντας τελεστές συναθροίσις για τους περιγραφείς των 2Δ περιοχών που σχηματίζουν τον όγκο. Ειδικά για τους περιγραφείς που βασίζονται στον υπολογισμό κάποιου ιστογράμματος, μπορούμε να επιλέξουμε τους τελεστές μέσου, ενδιάμεσου και τομής. Με αυτόν τον τρόπο δεν χρειάζεται να υπολογίσουμε εξ αρχής το ιστόγραμμα σε έναν 3Δ χώρο, αλλά με απλές πράξεις να επαναχρησιμοποιήσουμε τα ιστογράμματα που έχουν αποθηκευτεί για τις 2Δ περιοχές. Επιπρόσθετα, αποθηκεύουμε το μέγεθος και τη χωροχρονική θέση των όγκων με τη μορφή περικλειόντων πλαισίων (bounding box) για γρήγορο εντοπισμό και ανάκτηση.

4.3.3.3 3Δ Σημασιολογική Επέκταση

Η χωροχρονική κατάτμηση που βασίζεται σε οπτικά χαρακτηριστικά συνήθως δημιουργεί περισσότερους όγκους από τα αντικείμενα που υπάρχουν στην πραγματι-

κότητα μέσα στο βίντεο. Παρόμοια με τη λογική της σημασιολογικής επέκτασης περιοχών που αναλύσαμε στην ενότητα 4.2, επιχειρούμε να συγχωνεύσουμε όγκους με σημασιολογικά κριτήρια, όπως αυτά απεικονίζονται μέσα από το ασαφές σύνολο εννοιών του κάθε όγκου. Επιλέξαμε τον αλγόριθμο *SRSST* (υποενότητα 4.2.3) για τη σημασιολογική συγχώνευση όγκων, με τη διαφοροποίηση να έγκειται στη δημιουργία του σχεσιακού γράφου χαρακτηριστικών, καθώς αυτός πλέον αναπαριστά οντότητες στον 3Δ χώρο, όπως ορίστηκαν στην ενότητα 4.3.2.

Το αποτέλεσμα είναι να μειωθεί ο αριθμός των όγκων σε κάθε BoF και να έχουμε μία πιο συνεκτική κατάτμηση των αντικειμένων του βίντεο. Τέλος, τα σημασιολογικά χαρακτηριστικά του ανανεωμένου *ARG* που προκύπτει υπολογίζονται με τη σχέση (4.6), ενώ τα οπτικά χαρακτηριστικά υπολογίζονται ακολουθώντας την προσέγγιση που περιγράψαμε στην υποενότητα 4.3.3.2 και αυτός ο *ARG* θα χρησιμοποιηθεί στο επόμενο στάδιο της ανάλυσης διαδοχικών BoFs.

4.3.4 Ανάλυση Διαδοχικών BoFs

Στην προηγούμενη ενότητα (4.3.3) περιγράψαμε την κατάτμηση και την ταξινόμηση όγκων μέσα σε ένα μπλοκ από καρέ (BoF). Καθώς ο τελικός μας στόχος είναι η κατάτμηση ολόκληρης της ακολουθίας βίντεο που περιέχει συνεχείς και συνεκτικούς όγκους, οι οποίοι έχουν χρονική διάρκεια μεγαλύτερη του ενός BoF, χρειάζεται να εφαρμόσουμε μία στρατηγική για ταίριασμα όγκων και συνένωση διαδοχικών BoFs. Στις επόμενες δύο υποενότητες αναλύουμε την προτεινόμενη τεχνική ταίριασματος όγκων με χρήση οπτικών και σημασιολογικών χαρακτηριστικών. Η σημασιολογική ομαδοποίηση γίνεται καταρχάς σε όγκους οι οποίοι έχουν χαρακτηριστεί από κάποια έννοια με μεγάλο βαθμό συμμετοχής και επεκτείνεται τόσο χρονικά (σε διαδοχικά καρέ) όσο και χωρικά (σε γειτονικούς όγκους) βάσει της σημασιολογικής και της οπτικής τους συνάφειας.

4.3.4.1 Ταίριασμα Όγκων Διαδοχικών BoFs

Όπως έχουμε προαναφέρει, ένα BoF αποτελείται από ένα μικρό αριθμό από καρέ (συνήθως 10-15), οπότε είναι απαραίτητος ένας μηχανισμός συνένωσης διαδοχικών BoFs, έτσι ώστε οι εξαχθέντες όγκοι να έχουν την απαραίτητη χρονική συνέχεια, όπως αυτή καθορίζεται από την τροχιά και τη διάρκεια εμφάνισής σε όλο το πλάνο. Έστω δύο διαδοχικά BoFs B_i και B_{i+1} και οι αντίστοιχοι σχεσιακοί γράφοι χαρακτηριστικών τους G_i και G_{i+1} , αντιστοίχως. Η συνένωση των B_i και B_{i+1} θα γίνει με βάση το ποιοι όγκοι αυτών των δύο BoFs ταϊριάζουν και σε τι βαθμό. Οι πιθανοί συνδυασμοί είναι $|V(G_i)| \times |V(G_{i+1})|$ και σε αυτή την υποενότητα θα αναλύσουμε το πως μπορούμε να ελαττώσουμε αυτήν την πολυπλοκότητα βάσει ορισμένων κριτηρίων.

Καταρχάς, επειδή μελετάμε βίντεο ενός μόνο πλάνου, μπορούμε να κάνουμε την υπόθεση ότι τα αντικείμενα εμφανίζουν μία χωροχρονική συνέχεια και να περιορίσουμε τον αριθμό των πιθανών ταϊριασμάτων των όγκων, καθώς δεν είναι αναγκαίο να εξετάσουμε ακραίες περιπτώσεις χωροχρονικής ασυνέχειας. Έστω, G , ο γράφος που θα προκύψει από τη συγχώνευση των G_i και G_{i+1} . Αρχικά, θέτουμε $G = G_i \cup G_{i+1}$, ή ισοδύναμα $V(G) = V(G_i) \cup V(G_{i+1})$ και $E(G) = E(G_i) \cup E(G_{i+1})$. Για δύο κόμβους $v_a \in G_i$ και $v_b \in G_{i+1}$ θα δημιουργήσουμε μία (χρονική) ακμή στον G αν το περικλείον πλαίσιο του όγκου $b \in B_{i+1}$ τέμνεται με την κόλουμερη πυραμίδα που αναπαριστά τις πιθανές θέσεις του όγκου a στο επόμενο BoF B_{i+1} . Η πάνω βάση της πυραμίδας αντιστοιχεί ακριβώς στο περικλείον πλαίσιο του όγκου a , ενώ

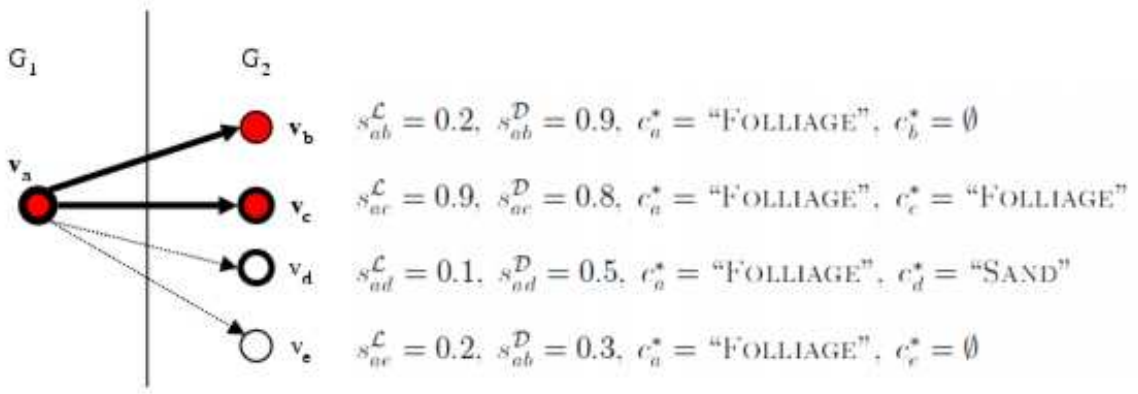
η μεγάλη (κάτω) βάση είναι η μεγέθυνση της πάνω βάσης κατά έναν παράγοντα $s = d_{max} \cdot h_{pyr}$, όπου d_{max} είναι η μέγιστη επιτρεπόμενη μετατόπιση ενός αντικειμένου μεταξύ δύο διαδοχικών καρέ, ενώ h_{pyr} είναι το ύψος της πυραμίδας κατά μήκος του χρονικού άξονα. Το σύνολο των ακμών μεταξύ των χωροχρονικά γειτονικών κόμβων το ορίζουμε ως $E_t(G)$, ή πιο απλά E_t . Το αποτέλεσμα είναι ότι για κάθε όγκο $a \in B_i$ υπάρχει ένα σύνολο όγκων πιθανών για ταίριασμα στο B_{i+1} , που μαθηματικά αναπαριστάται από το σύνολο των χρονικά γειτονικών όγκων N_a^t του κόμβου $v_a \in V(G_i)$: $N_a^t = \{v_b: v_b \in V(G_{i+1}), e_{ab} \in E_t(G)\}$. Με την ίδια διαδικασία σχηματίζουμε και το σύνολο N_b^t , όπου λόγω του τρόπου κατασκευής του (εξαιτίας της ασυμμετρίας της κόλυρης πυραμίδας) δεν ισχύει πάντα $N_a^t = N_b^t$.

Μετά το πρώτο βήμα, το οποίο δημιουργεί ένα σύνολο ακμών E_t μεταξύ χωροχρονικά γειτονικών όγκων που ανήκουν σε δύο διαδοχικά BoFs (βάση των χωροχρονικών περιορισμών που επιβάλλει η πυραμίδα), θέλουμε να περιορίσουμε περαιτέρω τους δυνατούς συνδυασμούς ταιριασμάτων. Από όλους τους όγκους επιλέγουμε να ξεκινήσουμε τον υπολογισμό των πιθανών ταιριασμάτων από αυτούς τους όγκους που χαρακτηρίζονται από μία κύρια έννοια (δες εξίσωση (4.3)), τους οποίους και ονομάζουμε κύριους όγκους και αντιστοιχούν στους κόμβους $v \in V_d(G_i) \subseteq V(G_i)$. Έστω $c_a^* \in \mathcal{L}_a$ και $c_b^* \in \mathcal{L}_b$ ότι είναι οι κύριες έννοιες για τους όγκους a και b , αντιστοίχως. Η διαδικασία ταιριάσματος όγκων ξεκινάει από τα ταιριάσματα δύο κύριων όγκων οι οποίοι χαρακτηρίζονται από την ίδια έννοια, δηλαδή για $v_a \in V_d(G_i)$, $v_b \in V_d(G_{i+1})$ και $c_a^* = c_b^*$. Το σύνολο των όγκων αυτών για το B_i είναι $V_d^*(G_i) \subseteq V_d(G_i) \subseteq V(G_i)$ και για το B_{i+1} αντίστοιχα $V_d^*(G_{i+1}) \subseteq V_d(G_{i+1}) \subseteq V(G_{i+1})$. Η ιδέα είναι ότι δύο χωροχρονικά γειτονικοί όγκοι a και b (σύμφωνα με τον ορισμό που δώσαμε στην παραπάνω παράγραφο), τους οποίους έχουμε χαρακτηρίσει με αξιοπιστία με την ίδια έννοια (σύμφωνα με τον ορισμό της κύριας έννοιας), αποτελούν το βέλτιστο δυνατό ταίριασμα. Το σύνολο των βέλτιστων δυνατών ταιριασμάτων αποτελεί το σύνολο των βέλτιστων ακμών $E_t^* \subseteq E_t$ (κριτήριο (α) της παράστασης (4.11)).

Το επόμενο βήμα είναι να εξετάσουμε τα πιθανά ταιριάσματα μεταξύ των κύριων όγκων από το B_i (έστω $V_d(G_i)$) και των χωροχρονικά γειτόνων αυτών στο B_{i+1} εξαιρουμένων των κύριων όγκων (έστω $\{N_{V_d(G_i)}^t - V_d(G_{i+1})\}$). Σε αυτή την περίπτωση λαμβάνουμε υπόψιν μας εκτός από τη σημασιολογική ομοιότητα $s_{ab}^{\mathcal{L}}$ και την οπτική ομοιότητα $s_{ab}^{\mathcal{D}}$ που έχουν οι δύο όγκοι a και b . Υπολογίζουμε τον κανονικοποιημένο βαθμό της οπτικής ομοιότητας $s_{ab}^{\mathcal{D}}$ με φθίνουσα σειρά και εφαρμόζουμε το κριτήριο (γ) της παράστασης (4.11) για να ελέγξουμε αν πρέπει να δημιουργήσουμε το ταίριασμα αυτό.

Τέλος, υπάρχει και η περίπτωση που οι δύο όγκοι a και b είναι κύριοι όγκοι, με διαφορετική όμως κύρια έννοια, $c_a^* \neq c_b^*$, γεγονός που μπορεί να οφείλεται στη δεδομένη αβεβαιότητα που έχει η διαδικασία της εξαγωγής όγκων και της ταξινόμησης αυτών. Τότε, με επιφύλαξη για την ορθότητα της ταξινόμησης του ενός από τους δύο όγκους, ελέγχουμε με ένα πολύ αυστηρό κριτήριο αν οι δύο όγκοι έχουν πάρα πολύ παρόμοια οπτική περιγραφή (κριτήριο (β) της παράστασης (4.11)). Το κριτήριο αυτό επιβάλλει η οπτική ομοιότητα $s_{ab}^{\mathcal{D}}$ να είναι η μεγαλύτερη μεταξύ όλων των πιθανών ταιριασμάτων ($rank(s_{ab}^{\mathcal{D}}) = 1$).

Τα παραπάνω κριτήρια για το ταίριασμα δύο όγκων συγκεντρώνονται στην παρα-



Σχήμα 4.9: Ταίριασμα όγκων διαδοχικών BoFs. Οι κύριοι όγκοι απεικονίζονται με παχιά κυκλική γραμμή, ενώ οι ακμές που τελικά δημιουργούνται είναι οι e_{ab} και e_{ac} που πληρούν τα κριτήρια (γ) και (α) της παράστασης (4.11) αντιστοίχως. Οι ακμές e_{ad} και e_{ae} (με διακεκομμένη γραμμή) εμπίπτουν στις περιπτώσεις (β) και (γ) χωρίς όμως να πληρούν τα αντίστοιχα κριτήρια.

κάτω παράσταση και αποτυπώνονται με ένα παράδειγμα στο σχήμα 4.9.

$$e_{ab} \in E_t(G) \text{ αν } \begin{cases} \{v_a \in V_d(G_i), v_b \in N_a^t \cap V_d(G_{i+1}), c_a^* = c_b^*\} \Leftrightarrow e_{ab} \in E_t^* & (\alpha) \\ \{v_a \in V_d(G_i), v_b \in N_a^t \cap V_d(G_{i+1}), c_a^* \neq c_b^*, \text{rank}(s_{ab}^{\mathcal{D}}) = 1\} & (\beta) \\ \{v_a \in V_d(G_i), v_b \in \{N_a^t - V_d(G_{i+1})\}, \frac{|E_a| - \text{rank}(s_{ab}^{\mathcal{D}})}{|E_a| - 1} \geq T_s\} & (\gamma) \end{cases} \quad (4.11)$$

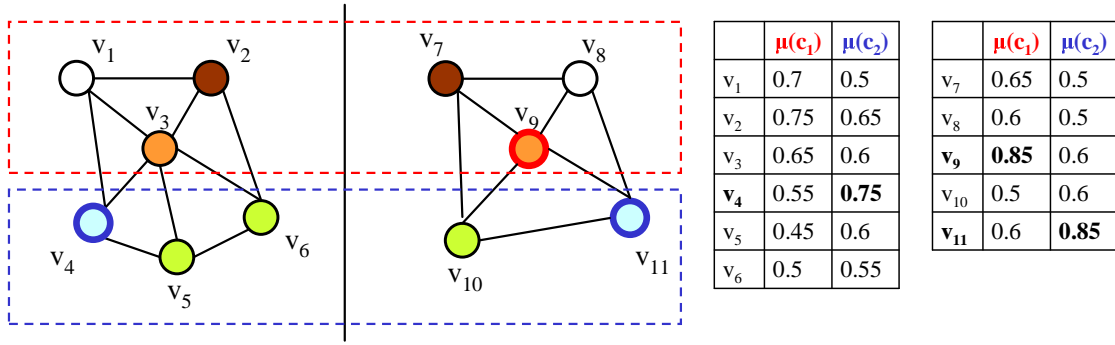
Το κατώφλι $T_s \in [0, 1]$ καθορίζει το πόσο ανεκτικό είναι το κριτήριο για την οπτική ομοιότητα. Στην ακραία περίπτωση που $T_s = 1$, τότε λαμβάνεται υπόψιν μόνο το καλύτερο οπτικό ταίριασμα (μέγιστο $s^{\mathcal{D}}$).

Στο σχήμα 4.9 φαίνεται ένα απλό παράδειγμα: Τα δύο διαδοχικά BoFs B_1 και B_2 , έχουν ενδεικτικά 1 και 4 κόμβους, δηλαδή $V(G_1) = \{v_a\}$ και $V(G_2) = \{v_b, v_c, v_d, v_e\}$. Οι όγκοι a και c, d είναι κύριοι όγκοι (παχιά γραμμή), δηλαδή $v_a \in V_d(G_1)$, $v_c, v_d \in V_d(G_2)$. Το πρώτο βήμα είναι να ελέγξουμε αν μπορούμε να ταιριάξουμε τους κύριους όγκους, δηλαδή αν μπορούμε να δημιουργήσουμε τις ακμές e_{ac} και e_{ad} . Για την ακμή e_{ac} αυτό είναι ορθό, καθώς $c_a^* = c_c^* = \text{FOLLIAGE}$ (κριτήριο (α)), ενώ για την e_{ad} δεν είναι ορθό καθώς $c_d^* = \text{SAND}$ και $\text{rank}(s_{ad}^{\mathcal{D}}) \neq 1$. Το δεύτερο βήμα είναι να ελέγξουμε αν μπορούμε να ταιριάξουμε τον όγκο a με έναν εκ των b και e , οι οποίοι δεν χαρακτηρίζονται από κάποια κύρια έννοια. Ελέγχουμε το (γ) κριτήριο και για την περίπτωση των a και b η οπτική ομοιότητα $s_{ab}^{\mathcal{D}}$ είναι αρκετά μεγάλη και δημιουργούμε την ακμή e_{ab} , κάτι που δεν ισχύει για την $s_{ab}^{\mathcal{D}}$.

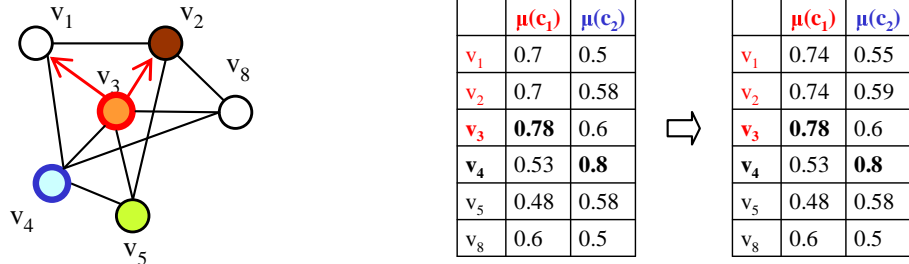
4.3.4.2 Ενημέρωση και Διάδοση Εννοιών

Εφόσον έχουν προσδιοριστεί ποιοι όγκοι ταιριάζουν μεταξύ δύο διαδοχικών BoFs, οι όγκοι αυτοί συγχωνεύονται και τα δύο BoFs συνενώνονται σε ένα. Επειδή η διαδικασία είναι επαναληπτική, θα εξεταστεί κατά πόσον το νέο BoF μπορεί να συνενωθεί με το επόμενο, ακολουθώντας τη μεθοδολογία που παρουσιάσαμε στην προηγούμενη ενότητα. Για να γίνει αυτό είναι απαραίτητο να ανανεώσουμε τα χαρακτηριστικά (οπτικά και σημασιολογικά) των νέων όγκων, όπως έχουν καθοριστεί μετά τις διαδικασίες ταιριάσματος και συγχώνευσης.

Η ανανέωση των εννοιών για κάθε όγκο γίνεται με χρήση ασαφών τελεστών συνάθροισης, οι οποίοι είναι κατάλληλοι, καθώς χρησιμοποιούμε ασαφή σύνολα για την



(α)



(β)

Σχήμα 4.10: Συγχώνευση δύο BoFs. (α) Ταίριασμα όγκων μεταξύ των δύο BoFs. (β) Συγχώνευση των δύο BoFs και ανανέωση των σημασιολογικών χαρακτηριστικών. Η ιδανική σημασιολογική κατάτμηση απεικονίζεται με τις διακεκομμένες γραμμές. Οι ταιριασμένοι όγκοι σημειώνονται με το ίδιο χρώμα και οι κύριοι όγκοι με παχιά γραμμή.

αναπαράσταση των σημασιολογικών χαρακτηριστικών. Οι κύριες έννοιες διαδίδονται μέσω των όγκων του νέου γράφου G . Έστω a ένας μη-κύριος όγκος, $v_a \in G$ ορίζουμε ένα σύνολο υποψήφιων κύριων εννοιών $C_a = \{c \in C | \mu_a(c) > T_c\}$. Για μία έννοια $c \in C_a$, υπολογίζουμε το βαθμό συμμετοχής $\hat{\mu}_a(c)$ που προκύπτει από τη συνάνθρωση του κόμβου v_a και των γειτονικών του που έχουν επίσης κύρια έννοια την c :

$$\hat{\mu}_a(c) = \frac{\sum_{b \in N_a^c} |b| \mu_b(c)}{\sum_{b \in N_a^c} |b|} \quad (4.12)$$

όπου $N_a^c = a \cup \{b \in N_a | c_b^* = c\}$ είναι η προαναφερθείσα γειτονιά και $|b|$ το τρέχον μέγεθος του όγκου b . Η έννοια $c^* \in C_a$ που μεγιστοποιεί το βαθμό $\hat{\mu}_a(c)$ επιλέγεται για κύρια έννοια του όγκου a και ανανεώνεται το ασαφές σύνολο των εννοιών \mathcal{L}_a με τελεστές ασαφούς συνάνθρωσης των όγκων της γειτονιάς $N_a^{c^*}$.

Η διάδοση αυτή εκτελείται σε όλο το γράφο G αναδρομικά. Έστω ο γράφος $G_D \equiv (V_d(G), E_{V_d(G)})$ ο οποίος είναι ένας υπογράφος του G που περιέχει μόνο τους κόμβους των κύριων όγκων ($V_d(G)$) και τις προσπίπτουσες ακμές αυτών ($E_{V_d(G)}$). Μόλις εξεταστούν και ανανεωθούν οι κύριοι όγκοι, είναι πιθανό να προκύψουν νέοι κύριοι όγκοι στον υπογράφο $G' = G - G_D$. Η αναδρομική ρουτίνα ανανέωσης των σημασιολογικών χαρακτηριστικών επαναλαμβάνεται στον υπογράφο G' , έως ότου να μην υπάρχουν νέοι κύριοι όγκοι, δηλαδή $G_D = \emptyset$. Το αποτέλεσμα είναι η διόρθωση των βαθμών συμμετοχής για τις έννοιες στους μη-κύριους όγκους, λαμβάνοντας υπόψη το εννοιολογικό πλαίσιο της γειτονιάς.

Το σχήμα 4.10 απεικονίζει ένα παράδειγμα για το πως επιτυγχάνεται η συγχώνευση όγκων με παράλληλη συνένωση των δύο BoFs. Στο παράδειγμα θεωρούμε ότι η ιδανική κατάτμηση απεικονίζεται με τις διακεκομμένες γραμμές, έχουμε δηλαδή δύο όγκους με κύριες έννοιες, τις c_1 και c_2 . Πριν τη συγχώνευση, έχουν εντοπιστεί ορισμένοι κύριοι όγκοι (v_4, v_9, v_{11}) στα δύο BoFs. Μετά τη συγχώνευση (σχήμα 4.10(β)) οι βαθμοί συμμετοχής ανανεώνονται σύμφωνα με την εξίσωση (4.6) και η σημασιολογική ομοιότητα s^c των νέων ακμών επαναυπολογίζεται σύμφωνα με την εξίσωση (4.8). Νέα στοιχεία αναδύονται για τη σημασιολογική ομοιότητα των ζευγών κόμβων (v_3, v_1) and (v_3, v_2), καθώς ο v_3 έχει αντιστοιχηθεί με τον κύριο όγκο v_9 . Επομένως, λόγω της διάδοσης της έννοιας c_1 , οι κόμβοι v_1 και v_2 συγχωνεύονται με τον v_3 και οι βαθμοί συμμετοχής τους στο ασαφές σύνολο των εννοιών ανανεώνονται βάσει της (4.12).

4.4 Οπτικό Εννοιολογικό Πλαίσιο

Η ιδέα να χρησιμοποιήσουμε οπτική εννοιολογική πληροφορία, στηρίζεται στο γεγονός ότι στη φύση κάθε αντικείμενο εντοπίζεται σε ένα σύνθητες περιβάλλον και συνυπάρχει μαζί με άλλα αντικείμενα. Το οπτικό εννοιολογικό πλαίσιο επιχειρεί να μοντελοποιήσει τις σχέσεις που διέπουν ορισμένες έννοιες, που απαντώνται σε εικόνες και βίντεο και να αξιοποιήσει αυτή την πληροφορία για τη βελτίωση των αποτελεσμάτων της ταξινόμησης περιοχών και όλης της εικόνας. Ορισμός και αναλυτική περιγραφή του εννοιολογικού πλαισίου και μερικών εφαρμογών σε προβλήματα ανάλυσης εικόνας και βίντεο μπορούν να αναζητηθούν στη διδακτορική διατριβή του Φ. Μυλωνά [85] και στην εκτενή κάλυψη της σχετικής βιβλιογραφίας που περιέχεται εκεί.

Σε αυτή την ενότητα το εννοιολογικό πλαίσιο το προσδιορίζουμε ως το βαθμό συσχέτισης μίας έννοιας με το θεματικό πεδίο στο οποίο ανήκει μία εικόνα. Παραδείγματος χάριν, η έννοια ΘΑΛΑΣΣΑ σχετίζεται με το θεματικό πεδίο “παραλία” σε βαθμό 0.9, ενώ η έννοια ΒΛΑΣΤΗΣΗ σε βαθμό 0.6. Όπως αναφέραμε στην ενότητα 2.4 οι έννοιες \mathcal{C} που ανήκουν σε μία οντολογία \mathcal{O} , συνδέονται μεταξύ τους με δυαδικές σχέσεις \mathcal{R} . Διατρέχουμε το γράφο της οντολογίας κάθε θεματικού πεδίου προκειμένου να υπολογίσουμε τη συσχέτιση μεταξύ δύο οποιονδήποτε εννοιών καθώς και τη σχετικότητα πλαισίου της κάθε έννοιας, με βάση τις σχέσεις \mathcal{R} και την ιδιότητα του μεταβατικού κλεισίματος. Οι βαθμοί συσχέτισης, επομένως, εξάγονται με στατιστική ανάλυση των δεδομένων εκπαίδευσης.

Η προτεινόμενη χρήση του οπτικού εννοιολογικού πλαισίου έχει ως σκοπό τη βελτίωση των εκτιμώμενων βαθμών συμμετοχής μίας περιοχής σε μία έννοια, όπως αυτοί προκύπτουν από τη σημασιολογική κατάτμηση μίας εικόνας. Συνοπτικά ο αλγόριθμος ακολουθεί τα παρακάτω βήματα:

1. Για κάθε έννοια $c \in \mathcal{C}$ που ανήκει στο ασαφές σύνολο εννοιών \mathcal{L}_a της περιοχής s_a με βαθμό συμμετοχής $\mu_a(c)$, υπολογίζουμε τη συσχέτιση $r_{cc'}$ με οποιαδήποτε άλλη έννοια $c' \neq c$ που ανήκει στο συγκεκριμένο θεματικό πεδίο θ .
2. Με χρήση μεταβατικού κλεισίματος στις σχέσεις $r_{cc'}$, $c' \in \mathcal{C} - \{c\}$, υπολογίζουμε το βαθμό συσχέτισης $cr_\theta(c)$ της έννοιας c με το θεματικό πεδίο θ .
3. Επαναυπολογίζουμε το βαθμό συμμετοχής $\mu_a(c)$ σύμφωνα με τον ακόλουθο αναδρομικό τύπο:

$$\mu_a^n(c) = \mu_a^{n-1}(c) - nr_\theta(\mu_a^{n-1}(c) - cr_\theta(c)) \quad (4.13)$$

Στην παραπάνω εξίσωση, n είναι ακέραιος αριθμός που υποδηλώνει τον αριθμό των επαναλήψεων που χρειάζονται για τη σύγκλιση στη βέλτιστη τιμή του $\mu_a^n(c)$ (συνήθως 3-4 επαναλήψεις αρκούν), ενώ np_{θ} είναι μία παράμετρος κανονικοποίησης η οποία εξαρτάται αποκλειστικά από το εκάστοτε θεματικό πεδίο θ και υποδηλώνει τη σημασιολογική συσχέτιση των εννοιών του πεδίου.

Η συνεισφορά του οπτικού πλαισίου στη σημασιολογική κατάτμηση έγκειται στην αναπροσαρμογή των βαθμών συμμετοχής των εννοιών, ανάλογα με το πόσο σχετικές είναι στο θεματικό πεδίο, συνυπολογίζοντας παράλληλα και τη συσχέτιση των υπόλοιπων εντοπισμένων εννοιών στην εικόνα. Όπως θα φανεί και στην επόμενη ενότητα, τα πειραματικά αποτελέσματα υποδεικνύουν ότι το εννοιολογικό πλαίσιο ενισχύει τα ορθά αποτελέσματα της αρχικής αναγνώρισης εννοιών και ελαττώνει τους βαθμούς συμμετοχής των εσφαλμένων εννοιών, μειώνοντας την επίδραση που έχουν οι περιοχές αυτές στη σημασιολογική κατάτμηση.

4.5 Πειραματικά Αποτελέσματα σε Εικόνες

4.5.1 Προετοιμασία Πειραμάτων και Διαδικασία Αξιολόγησης

Προκειμένου να αξιολογήσουμε τον αλγόριθμο σημασιολογικής κατάτμησης, στις δύο παραλλαγές του, κάναμε πειράματα σε δύο σύνολα εικόνων που ανήκουν στα θεματικά πεδία “παράλια” και “ράλι”. Το σύνολο των εικόνων του πρώτου θεματικού πεδίου ήταν 443 και του δεύτερου 159, οι οποίες αποκτήθηκαν είτε από το διαδίκτυο είτε από προσωπική συλλογή. Από τις συνολικά 602 εικόνες χρησιμοποιήσαμε το 20% αυτών, ήτοι 120 εικόνες, για την εκπαίδευση και τον υπολογισμό των παραμέτρων T_{seed} , T_{merge} , T_w και np . Είναι κοινώς αποδεκτό [146, 73] ότι η πιο αντικειμενική μεθοδολογία αξιολόγησης αλγορίθμων κατάτμησης περιλαμβάνει μία σχετική σύγκριση των αποτελεσμάτων σε σχέση με αυτά που θεωρούνται ορθά, το σύνολο δεδομένων αλήθειας (ground truth). Η κατασκευή του συνόλου δεδομένων αλήθειας είναι μία χρονοβόρα διαδικασία, την οποία προσπαθήσαμε να απλοποιήσουμε με την υλοποίηση ενός εργαλείου χειροκίνητης κατάτμησης και ταξινόμησης περιοχών. Προκειμένου να παρουσιάσουμε βήμα προς βήμα τη μεθοδολογία και τα αποτελέσματά της, υλοποιήσαμε ένα σύστημα γραφικής αλληλεπίδρασης με το χρήστη, το οποίο ενσωματώνει όλους τους προτεινόμενους αλγορίθμους όσο και τη διαδικασία της αξιολόγησης αυτών.

Η διαδικασία αξιολόγησης είναι πάντα ιδιαίτερα σημαντική καθώς ποσοτικοποιεί την επίδοση ενός αλγορίθμου ή ενός συστήματος, βοηθώντας στην εξαγωγή επιστημονικών και συνεπών συμπερασμάτων. Καθώς στο κεφάλαιο αυτό έχουμε χρησιμοποιήσει ασαφή σύνολα για την αναπαράσταση των αποτελεσμάτων της ταξινόμησης, θεωρήσαμε φυσικό να υιοθετήσουμε τη χρήση ασαφών τελεστών για την αξιολόγηση των αποτελεσμάτων. Τα τελικά αποτελέσματα των δύο αλγορίθμων σημασιολογικής κατάτμησης, όσο και της εφαρμογής του αλγορίθμου εννοιολογικού πλαισίου, είναι η μάσκα κατάτμησης και ένα αρχείο που αναπαριστά το ασαφές σύνολο εννοιών για όλες τις περιοχές της κατατμημένης εικόνας. Η εικόνα από το σύνολο δεδομένων αλήθειας αποτελείται από ένα σύνολο συνδεδεμένων, μη επικαλυπτόμενων περιοχών seg_i , οι οποίες είναι συνδεδεμένες χειρωνακτικά με μία αποκλειστικά έννοια.

Αρχικά υπολογίζουμε το ποσοστό της επικάλυψης της κάθε περιοχής a με όλες τις περιοχές seg_i της μάσκας από την εικόνα του συνόλου δεδομένων αλήθειας, βάσει

της εξίσωσης (4.14), όπου η ποσότητα $|a|$ είναι το μέγεθος της περιοχής a , δηλαδή το εμβαδόν της μετρημένο σε εικονοστοιχεία.

$$\text{overlap}(a, \text{seg}_i) = \frac{|a \cap \text{seg}_i|}{|a|} \quad (4.14)$$

Στη συνέχεια υπολογίζουμε την Dombi τ-νόρμα [15] (η οποία επιλέχθηκε μεταξύ άλλων τ-νορμών μετά από πειραματικό έλεγχο) μεταξύ της επικάλυψης και του βαθμού συμμετοχής της αντίστοιχης έννοιας που ορίζει το σύνολο δεδομένων αλήθειας για την εν λόγω περιοχή. Το αποτέλεσμα του τελεστή αυτού μετράει την επιτυχή κατάτμηση και αναγνώριση της συγκεκριμένης περιοχής σε σχέση και με την αναγνώριση της ορθής (ως προς το σύνολο δεδομένων αλήθειας) έννοιας c_i .

$$\text{score}_a(c_i) = \max_i (t_{\text{dombi}}(\text{overlap}(a, \text{seg}_i), \mu_a(c_i))) \quad (4.15)$$

Χρησιμοποιούμε την εξίσωση (4.15) για όλες τις έννοιες του ασαφούς συνόλου A της περιοχής a και τις συνδυάζουμε για να υπολογίσουμε την επίδοση του αλγορίθμου στη συγκεκριμένη περιοχή, όπου u_{dombi} είναι η Dombi τ-κονόρμα:

$$\text{score}_a = u_{\text{dombi}}(\{\text{score}_a(c_i) : c_i \in A\}) \quad (4.16)$$

Χάρη στην ιδιότητα της αντιμεταθετικότητας των τελεστών τ-κονόρμας των ασαφών συνόλων, η σειρά των ορισμάτων της (4.16) δεν παίζει κάποιο σημαντικό ρόλο, οπότε ο υπολογισμός γίνεται σειριακά για ζεύγη βαθμών συμμετοχής των εννοιών $c_i \in A$. Η εξίσωση (4.16) μας παρέχει ένα μέτρο αξιολόγησης της επίδοσης του αλγορίθμου για μία συγκεκριμένη περιοχή, ανεξάρτητα της έννοιας που απεικονίζει. Αν και πρόκειται για μία ενδιαφέρουσα ποσότητα, κυρίως αποτελεί μία ενδιαμέση τιμή για να υπολογίσουμε την επίδοση σε όλη την εικόνα. Καθώς η κατατμημένη εικόνα αποτελείται από την ένωση των επιμέρους τμημάτων, βασιζόμενοι στο score_a μπορούμε να αξιολογήσουμε το αποτέλεσμα του αλγορίθμου σε όλη την εικόνα, χρησιμοποιώντας τελεστές συναθροίσεως:

$$\text{score} = \frac{\sum_{a \in I} (|a| \cdot \text{score}_a)}{\sum_{a \in I} |a|} \quad (4.17)$$

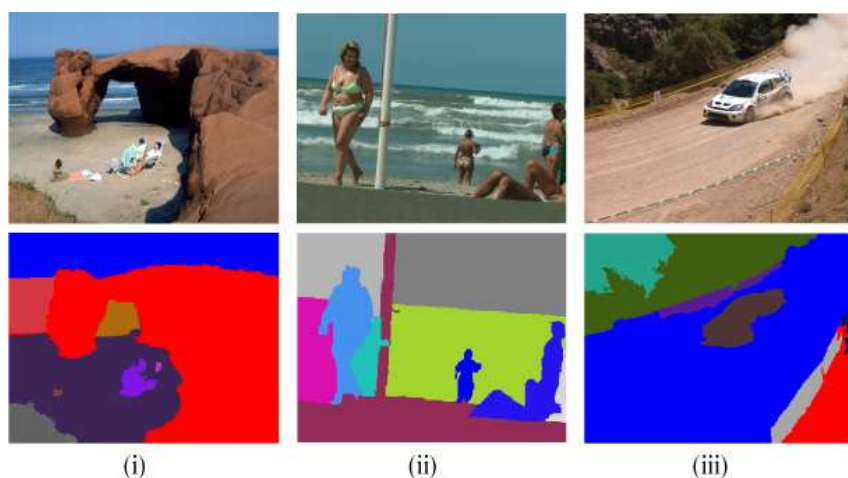
Η παραπάνω μεθοδολογία αξιολόγησης, με χρήση των εξισώσεων (4.14), (4.15), (4.16) και (4.17) παρέχει τη δυνατότητα αξιολόγησης τόσο των αλγορίθμων κατάτμησης, όσο και της επίδοσης στην αναγνώριση των εννοιών και στον εντοπισμό της εμφάνισής τους μέσα στην εικόνα. Η τιμή $\text{score} \in [0, 1]$ για την κάθε εικόνα αντανάκλα την επίδοση του αλγορίθμου σημασιολογικής κατάτμησης, λαμβάνοντας υπόψιν την ποιότητα της κατάτμησης και τον εντοπισμό των ορθών εννοιών στις αντίστοιχες περιοχές, πάντα σε σχέση με το σύνολο δεδομένων αλήθειας. Προκειμένου να συγκρίνουμε τα αποτελέσματα των προτεινόμενων αλγορίθμων σημασιολογικής κατάτμησης με αυτά ενός παραδοσιακού αλγόριθμου, τα οποία προφανώς αγνοούν οποιαδήποτε σημασιολογική πληροφορία για τις περιοχές της κατατμημένης εικόνας, χρησιμοποιούμε τις μάσκες κατάτμησης ως είσοδο σε έναν αλγόριθμο κατάταξης, ο οποίος στηρίζεται σε ταίριασμα οπτικών περιγραφών και εξαγωγή εννοιών ανά περιοχή, όπως περιγράφεται στο κεφάλαιο 3. Στη συνέχεια του κεφαλαίου θα δείξουμε ότι η διαφορά της χρήσης της σημασιολογικής πληροφορίας που περιέχουν τα ασαφή σύνολα εννοιών

της κάθε περιοχής κατά τη διάρκεια της κατάτμησης, οδηγούν στη βελτίωση τόσο της κατάτμησης όσο και της αναγνώρισης και του εντοπισμού των εννοιών σε κάθε περιοχή.

Τα αποτελέσματα της αξιολόγησης που έγιναν σε όλο το σύνολο των εικόνων είναι θετικά, ακόμα και σε περιπτώσεις οι οποίες χαρακτηρίζονται από δυσκολίες στην ορθή κατάτμηση και στον εντοπισμό εννοιών. Η επέκταση περιοχών καθοδηγείται και από τις πιθανές έννοιες που εντοπίζονται, οδηγώντας σε βελτίωση των αποτελεσμάτων των αντίστοιχων παραδοσιακών αλγορίθμων. Τέλος, οι περιοχές της εικόνας που προκύπτουν από τη σημασιολογική κατάτμηση αντιστοιχίζονται με τις σημασιολογικές έννοιες, τις οποίες αναπαριστούν με ένα βαθμό που προσδίδει ένα λογικό μέτρο πίστης για την ορθή σημασιολογική κατάτμηση.

4.5.2 Ενδεικτικά Αποτελέσματα και Συμπεράσματα

Σε αυτή την παράγραφο θα περιγράψουμε λεπτομερώς τρία πειραματικά παραδείγματα, προκειμένου να παρουσιάσουμε καλύτερα τις προτεινόμενες μεθοδολογίες σημασιολογικής κατάτμησης: τα παραδείγματα προέρχονται από τα θεματικά πεδία “παραλία” και “ράλι”. Συμπληρωματικά δίνουμε και ένα παράδειγμα υπολογισμού της σημασιολογικής ομοιότητας μεταξύ δύο γειτονικών περιοχών, καθώς και την επίδραση του αλγορίθμου του εννοιολογικού πλαισίου σε δύο εικόνες στα αποτελέσματα της ταξινόμησης. Για τη συνολική επισκόπηση και εκτίμηση της επίδοσης των προτεινόμενων μεθοδολογιών, παρουσιάζουμε τα αποτελέσματα για το σύνολο των εικόνων, των θεματικών πεδίων και των αλγορίθμων.



Σχήμα 4.11: Εικόνες από το σύνολο δεδομένων αλήθειας (ground truth images).

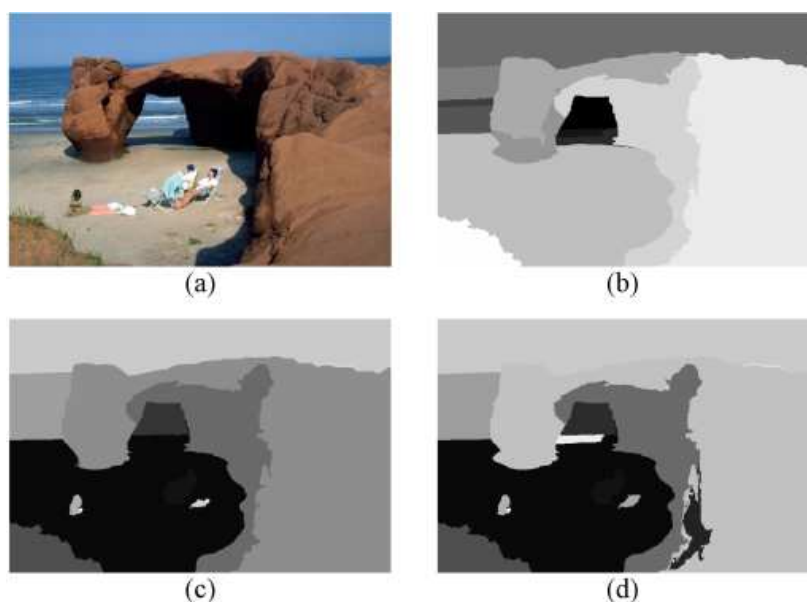
Κάθε ένα από τα τρία σύνολα εικόνων αποτελούνται από τέσσερις εικόνες: (α) την αρχική εικόνα, (β) το αποτέλεσμα του παραδοσιακού RSST αλγορίθμου κατάτμησης, (γ) τη μάσκα κατάτμησης του *SRSST* και τέλος (δ) τη μάσκα κατάτμησης του *SW*. Στο σχήμα 4.11 φαίνονται οι μάσκες κατάτμησης από το σύνολο δεδομένων αλήθειας, οι οποίες δημιουργήθηκαν χειρωνακτικά. Στην περίπτωση του RSST προκαθορίζουμε τον αριθμό των τελικών περιοχών ώστε να ισούται με αυτές που παράγονται από τον *SRSST* και να έχουμε συγκρίσιμα αποτελέσματα.

Στο σχήμα 4.12 φαίνονται τα αποτελέσματα για μία εικόνα από το θεματικό πεδίο “παραλία”, στην οποία εμφανίζονται άνθρωποι ξαπλωμένοι σε μία αμμώδη πα-

ραλία περιστοιχισμένη από βράχια. Σε μία πρώτη επισκόπηση, μπορούμε να διαπιστώσουμε ότι το αποτέλεσμα του RSST στο σχήμα 4.12 (a) είναι ανεπαρκές: οι άνθρωποι έχουν συγχωνευτεί με την περιοχή της άμμου και επιπλέον η θάλασσα και τα βράχια έχουν υπερκατατμηθεί σε πολλές περιοχές. Το αποτέλεσμα της χρήσης του *SW* (σχήμα 4.12 (c)) παρουσιάζει μία αισθητή βελτίωση. Πιο συγκεκριμένα παρατηρούμε ότι τα δύο σώματα της θάλασσας (στα αριστερά της εικόνας και στο μέσο ανάμεσα στα βράχια) έχουν κατατμηθεί σωστά, παρά τη διαφορετική υφή και χρώμα που παρουσιάζουν ελέω των κυματισμών της θάλασσας, παράγοντες που οδήγησαν στην εσφαλμένη υπερκατάτμηση στην περίπτωση του RSST. Επίσης, τα βράχια έχουν κατατμηθεί καλύτερα σε δύο μεγάλες περιοχές που δυστυχώς λόγω της έντονης παρουσίας σκιών στη μία εξ αυτών, δεν κατέστη δυνατό να της αντιστοιχηθεί η έννοια ΒΡΑΧΟΣ σε αρκετά μεγάλο βαθμό ώστε να συγχωνευτεί με τη γειτονική περιοχή που έχει ταξινομηθεί ως ΒΡΑΧΟΣ.

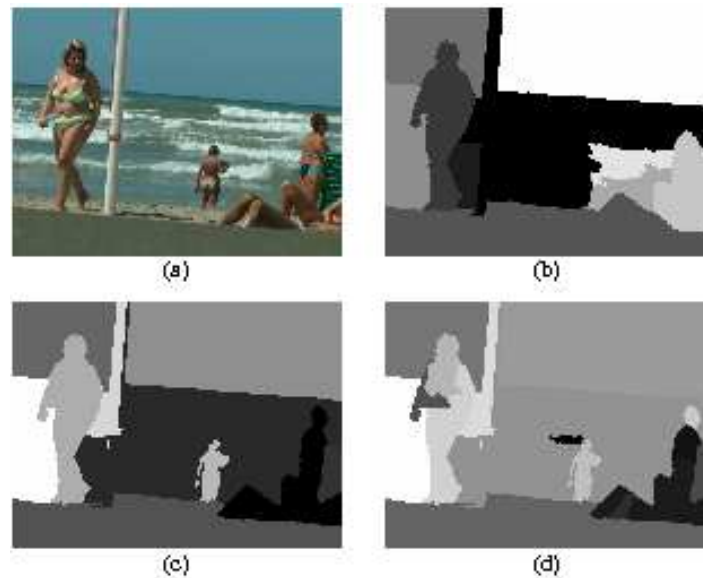
Ακόμη η σημασιολογική κατάτμηση *SW* καταφέρνει να επιτύχει σε έναν ικανοποιητικό βαθμό, την αναγνώριση και τον εντοπισμό των ανθρώπων που είναι ξαπλωμένοι στην άμμο. Καλά αποτελέσματα επιτεύχθηκαν και στην περίπτωση της έντονης σκιάς στο μέσο της εικόνας, καθώς η περιοχή κάτω από τα βράχια αναγνωρίστηκε σωστά ως ΑΜΜΟΣ, σε αντίθεση με τη λανθασμένη κατάτμηση στην περίπτωση του RSST.

Στη συνέχεια στο σχήμα 4.12 (d) φαίνονται τα αποτελέσματα της σημασιολογικής κατάτμησης *SRSST* για την ίδια εικόνα. Σε σχέση με τον αλγόριθμο *SW* (σχήμα 4.12 (c)) παρατηρούμε μικρές διαφορές. Για παράδειγμα, ο βράχος στα δεξιά αποτελείται από τρία τμήματα, ενώ και η θάλασσα κάτω από τα βράχια στο μέσο της εικόνας έχει λανθασμένα υπερκατατμηθεί. Για τη συγκεκριμένη εικόνα οι τιμές του *score*, όπως ορίστηκε στην προηγούμενη ενότητα, για κάθε μία από τις τρεις περιπτώσεις είναι 0.61 για τον RSST, 0.85 για τον *SW* και 0.78 για τον *SRSST*.



Σχήμα 4.12: Κατάτμηση εικόνας με χρήση των αλγορίθμων *RSST*, *SW* και *SRSST* (1ο παράδειγμα).

Σε αυτό το σημείο ας εξετάσουμε λεπτομερώς ένα βήμα του αλγορίθμου *SRSST*. Η αρχική κατάτμηση και ταξινόμηση παρήγαγε περίπου τριάντα περιοχές και ένα ασα-



Σχήμα 4.13: Κατάτμηση εικόνας με χρήση των αλγορίθμων *RSST*, *SW* και *SRSST* (2ο παράδειγμα).

Πίνακας 4.1: Βαθμοί συμμετοχής κάθε έννοιας για τέσσερις γειτονικές περιοχές.

Περιοχή	Έννοιες					
	ΟΥΡΑΝΟΣ	ΘΑΛΑΣΣΑ	ΒΡΑΧΟΣ	ΒΛΑΣΤΗΣΗ	ΑΜΜΟΣ	ΑΝΘΡΩΠΟΣ
a	0.66	0.82	0.67	0.65	0.68	0.75
b	0.74	0.77	0.64	0.64	0.69	0.73
c	0.68	0.79	0.68	0.65	0.75	0.67
d	0.90	0.77	0.67	0.65	0.64	0.98

φές σύνολο εννοιών για την κάθε μία. Σύμφωνα με το σύνολο δεδομένων αλήθειας, τέσσερις από αυτές (έστω περιοχές a, b, c, d) στην πραγματικότητα αντιστοιχούν σε μία περιοχή η οποία είναι η θάλασσα στα αριστερά της εικόνας του σχήματος 4.12 (a). Οι τέσσερις αυτές περιοχές σχηματίζουν έναν υπογράφο G_{sea} του *ARG*, δηλαδή $G_{sea} \equiv (V_{sea}, E_{sea})$, όπου $V_{sea} = v_a, v_b, v_c, v_d$ και $E_{sea} = e_{ab}, e_{ac}, e_{ad}, e_{cd}$.

Στον πίνακα 4.1 εμφανίζονται οι έννοιες και οι βαθμοί συμμετοχής τους για τις υπό εξέταση περιοχές a, b, c, d . Με βάση τις τιμές αυτές και την εξίσωση (4.5) υπολογίζουμε τα βάρη των ακμών του υπογράφου G_{sea} (πίνακας 4.2). Εύκολα παρατηρούμε ότι η ακμή με το μικρότερο βάρος είναι η e_{ac} , με $w(e_{ac}) = 0.21$, καθώς οι δύο αυτές περιοχές μοιράζονται την έννοια ΘΑΛΑΣΣΑ με βαθμό ομοιότητας 0.79. Σύμφωνα με τον αλγόριθμο *SRSST* οι περιοχές a, c θα συγχωνευτούν και θα δημιουργήσουν μία νέα περιοχή, έστω \hat{a} . Σύμφωνα με την εξίσωση (4.6), το ασαφές σύνολο εννοιών της νέας περιοχής υπολογίζεται με απλή αντικατάσταση των τιμών:

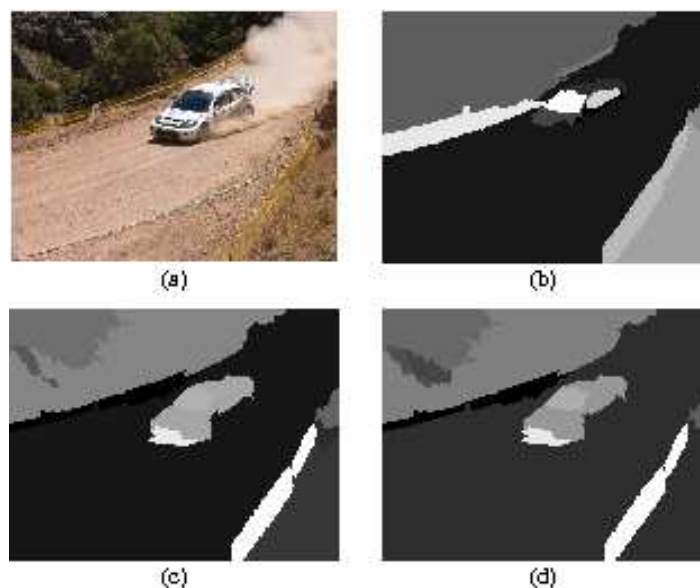
$$\mu_{\hat{a}}(\Theta_{\text{ΘΑΛΑΣΣΑ}}) = \frac{|a| \cdot \mu_a(\Theta_{\text{ΘΑΛΑΣΣΑ}}) + |c| \cdot \mu_c(\Theta_{\text{ΘΑΛΑΣΣΑ}})}{|a| + |c|} \approx 0.811$$

Με τον ίδιο τρόπο υπολογίζονται και τα $\mu_{\hat{a}}(\text{ΟΥΡΑΝΟΣ}), \mu_{\hat{a}}(\text{ΒΡΑΧΟΣ}), \dots$. Συνεχίζοντας στο επόμενο βήμα του αλγορίθμου, η ακμή e_{ac} αφαιρείται από το γράφο G_{sea} και για όλες τις γειτονικές ακμές (που έχουν έναν από τους δύο κόμβους τους, τον a ή τον c) επαναυπολογίζονται τα βάρη τους σύμφωνα με την (4.6) και το νέο ασαφές σύνολο εννοιών $L_{\hat{a}}$ της περιοχής \hat{a} .

Πίνακας 4.2: Υπολογισμός βαρών των ακμών του υπογράφου G_{SEA} .

Ακμές	Έννοιες						Βάρος w
	ΟΥΡΑΝΟΣ	ΘΑΛΑΣΣΑ	ΒΡΑΧΟΣ	ΒΛΑΣΤΗΣΗ	ΑΜΜΟΣ	ΑΝΘΡΩΠΟΣ	
e_{ab}	0.66	0.77	0.64	0.64	0.68	0.73	0.23
e_{ac}	0.66	0.79	0.67	0.65	0.68	0.67	0.21
e_{ad}	0.66	0.77	0.67	0.65	0.64	0.75	0.23
e_{cd}	0.68	0.77	0.64	0.64	0.69	0.67	0.23

Στο δεύτερο παράδειγμα (σχήμα 4.13), πάλι τα αποτελέσματα του RSST κρίνονται ανεπαρκή: ορισμένες περιοχές που περιέχουν ανθρώπους ή μέρη τους, είναι λανθασμένα συγχωνευμένες με περιοχές που περιγράφουν τη θάλασσα ή την άμμο, ενώ επιπλέον η θάλασσα είναι υπερκατατμημένη λόγω των χρωματικών αντιθέσεων των επιμέρους περιοχών. Η εφαρμογή του SW βελτιώνει σημαντικά τα αποτελέσματα (σχήμα 4.13 (c)). Η θάλασσα κατατμηνεται και εντοπίζεται σωστά, καθώς και η γυναίκα στο προσκήνιο. Το μειονέκτημα είναι η συγχώνευση των ανθρώπων στα δεξιά, καθώς ο αλγόριθμος δεν αντιμετωπίζει περιπτώσεις επικάλυψης αντικειμένων που ανήκουν στην ίδια έννοια. Για την ίδια εικόνα ο $SRSST$ (σχήμα 4.13 (d)) τα πηγαίνει εξίσου καλά με τον SW . Μειονεκτεί όπως φαίνεται στην περίπτωση της κατάτμησης της γυναίκας του προσκηνίου, καθώς ένα τμήμα έχει ταξινομηθεί λάθος και επηρεάζει την περαιτέρω συγχώνευσή του με το υπόλοιπο σώμα. Σε ποσοτικά νούμερα, για τη συγκεκριμένη εικόνα ο RSST πέτυχε $score = 0.82$, ο SW 0.90 και ο $SRSST$ 0.88.

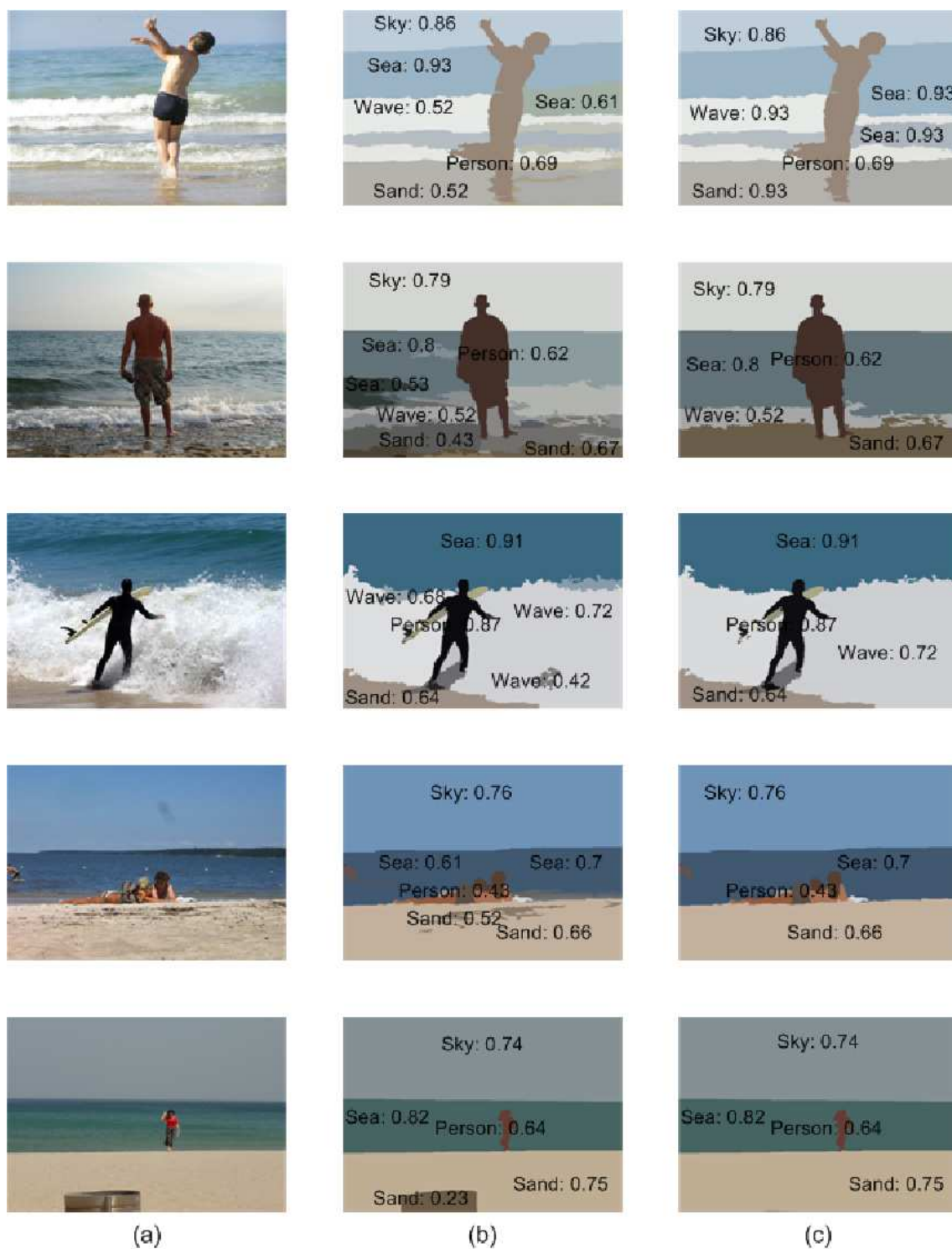


Σχήμα 4.14: Κατάτμηση εικόνας με χρήση των αλγορίθμων $RSST$, SW και $SRSST$ (3ο παράδειγμα).

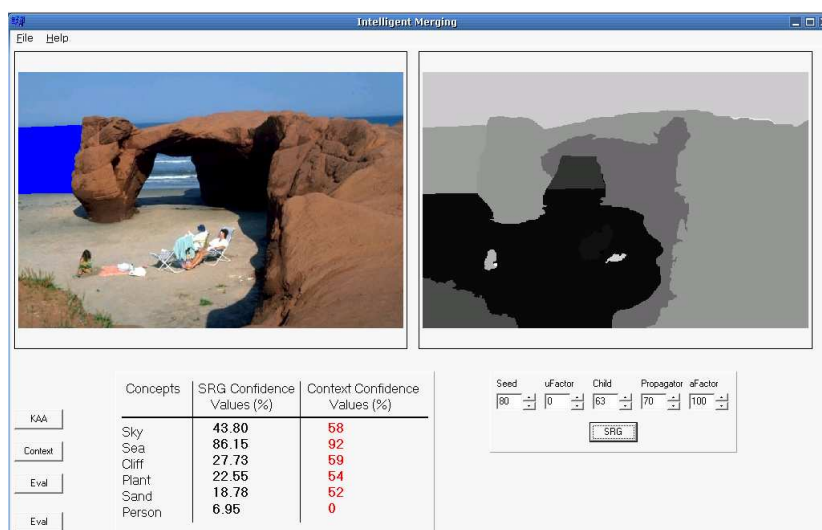
Αποτελέσματα από το θεματικό πεδίο “ράλι” εμφανίζονται στο σχήμα 4.14, (εμφανίζεται ένα στιγμιότυπο από αγώνα του παγκοσμίου πρωταθλήματος ράλι), όπου μπορούν να εξαχθούν παρόμοια συμπεράσματα με τα παραπάνω.

4.5.3 Εννοιολογικό Πλαίσιο

Στο σχήμα 4.16 παρουσιάζεται η γραφική διεπαφή χρήστη του συστήματος, και πιο συγκεκριμένα στη φάση ελέγχου των αποτελεσμάτων του αλγορίθμου του εννοιολο-



Σχήμα 4.15: Σημασιολογική κατάτμηση εικόνων και αναπροσαρμογή των βαθμών εμπιστοσύνης των εντοπισμένων εννοιών μέσω ανατροφοδότησης της αναγνώρισης.



Σχήμα 4.16: Σημασιολογική κατάτμηση και εννοιολογικό πλαίσιο.

Πίνακας 4.3: Βαθμοί συμμετοχής κάθε έννοιας πριν και μετά την εφαρμογή του αλγορίθμου εννοιολογικού πλαισίου για εικόνα από το θεματικό πεδίο “παραλία”.

Περιοχή	Έννοιες											
	Ουρανός		Θάλασσα		Βράχος		Βλάστηση		Άμμος		Άνθρωπος	
	πριν	μετά	πριν	μετά	πριν	μετά	πριν	μετά	πριν	μετά	πριν	μετά
s_0	0.66	0.72	0.82	0.90	0.67	0.73	0.65	0.69	0.68	0.70	0.75	0.74
s_3	0.73	0.75	0.72	0.85	0.64	0.72	0.66	0.69	0.64	0.69	0.73	0.75
s_7	0.69	0.73	0.78	0.88	0.66	0.72	0.65	0.69	0.72	0.72	0.93	0.81
s_9	0.69	0.73	0.66	0.83	0.64	0.72	0.67	0.70	0.66	0.69	0.65	0.69
s_{12}	0.66	0.71	0.70	0.85	0.67	0.73	0.65	0.69	0.74	0.73	0.81	0.91
s_{13}	0.64	0.71	0.71	0.85	0.69	0.74	0.64	0.69	0.73	0.75	0.75	0.73
s_{15}	0.90	0.82	0.77	0.88	0.67	0.73	0.65	0.69	0.64	0.69	0.98	0.84
s_{16}	0.79	0.74	0.80	0.89	0.64	0.72	0.66	0.69	0.65	0.69	0.73	0.75
s_{18}	0.84	0.95	0.83	0.90	0.67	0.73	0.89	0.80	0.65	0.69	0.00	0.40
s_{26}	0.69	0.73	0.90	0.93	0.69	0.74	0.66	0.69	0.65	0.69	0.84	0.77
s_{29}	0.76	0.76	0.61	0.71	0.68	0.74	0.74	0.73	0.70	0.75	0.87	0.79

γικού πλαισίου σε μία εικόνα. Ο αλγόριθμος αυτός εφαρμόζεται μετά το πέρας της σημασιολογικής κατάτμησης και όπως περιγράψαμε στην ενότητα 4.4 ρυθμίζει τους βαθμούς συμμετοχής για κάθε έννοια κάθε περιοχής χωρίς να μεταβάλλει τη μάσκα κατάτμησης της εικόνας. Στο σχήμα 4.16 φαίνεται ότι έχουμε επιλέξει την περιοχή της θάλασσας στο πάνω αριστερά σημείο της εικόνας (χρωματισμένο με έντονο μπλε). Κάτω από την εικόνα φαίνονται οι βαθμοί συμμετοχής κάθε έννοιας για την επιλεγμένη περιοχή, πριν και μετά την εφαρμογή του αλγορίθμου εννοιολογικού πλαισίου. Για τη συγκεκριμένη περιοχή, η οποία προφανώς αντιστοιχεί στην έννοια ΘΑΛΑΣΣΑ, ο αλγόριθμος αυξάνει την εν λόγω τιμή συμμετοχής από 0.86 σε 0.92. Για τις υπόλοιπες έννοιες, οι οποίες είναι άσχετες με τη συγκεκριμένη περιοχή, οι βαθμοί συμμετοχής τους είτε μειώνονται είτε αυξάνονται λίγο, διευκολύνοντας τη διάκριση της κυρίαρχης (και σωστής) έννοιας ΘΑΛΑΣΣΑ.

Για να δείξουμε καλύτερα την επίδραση του εννοιολογικού πλαισίου, παρουσιάζουμε στον πίνακα 4.3 τους βαθμούς συμμετοχής των έξι εννοιών ΟΥΡΑΝΟΣ, ΘΑΛΑΣΣΑ, ΒΡΑΧΟΣ, ΒΛΑΣΤΗΣΗ, ΑΜΜΟΣ, ΑΝΘΡΩΠΟΣ, για τις περιοχές που προέκυψαν μετά τη σημασιολογική κατάτμηση και μετά τη χρήση του εννοιολογικού πλαισίου (στήλες:

Πίνακας 4.4: Σύγκριση επιδόσεων ανά έννοια των προτεινόμενων αλγορίθμων για το θεματικό πεδίο “παραλία”.

Έννοιες	Watershed	SW	$SW + C$	RSST	$SRSST$	$SRSST + C$
ΟΥΡΑΝΟΣ	0.93	0.95	0.94	0.90	0.93	0.93
ΘΑΛΑΣΣΑ	0.79	0.80	0.84	0.81	0.83	0.89
ΑΜΜΟΣ	0.72	0.79	0.81	0.72	0.82	0.88
ΑΝΘΡΩΠΟΣ	0.44	0.56	0.60	0.48	0.48	0.55
ΒΡΑΧΟΣ	0.61	0.75	0.77	0.66	0.76	0.78
ΒΛΑΣΤΗΣΗ	0.67	0.70	0.71	0.63	0.68	0.69
Σύνολο	0.69	0.77	0.79	0.70	0.77	0.78

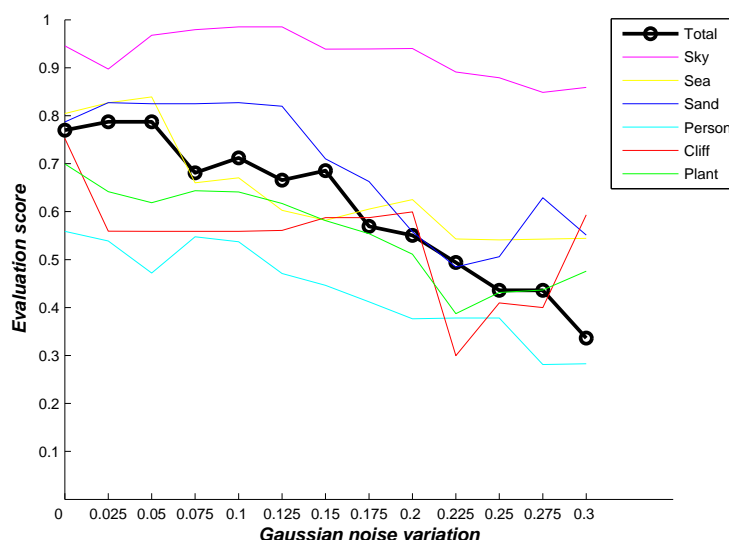
πριν - μετά) και συνοψίζουμε την επίδραση στα συνολικά αποτελέσματα της εικόνας. Η πραγματική έννοια που αντιστοιχεί σε κάθε περιοχή, όπως την έχουμε καθορίσει χειρωνακτικά κατά την κατασκευή του συνόλου δεδομένων αλήθειας, τονίζεται στον πίνακα με σκούρα γράμματα, έτσι ώστε να φανεί καλύτερα ότι το εννοιολογικό πλαίσιο βελτιώνει το βαθμό εμπιστοσύνης της αναγνώρισης των σωστών εννοιών. Ποσοτικά, οι περιοχές s_0, s_{18} αυξάνουν το βαθμό συμμετοχής της ορθής τους έννοιας σε ποσοστό περίπου 11% (για την έννοια ΘΑΛΑΣΣΑ) και 13% (ΟΥΡΑΝΟΣ) αντίστοιχα. Συνολικά, μετρήσαμε μία αύξηση της τάξης του 12.5% για τις σωστές έννοιες κάθε περιοχής, ποσοστό διόλου αμελητέο.

4.5.4 Συνολικά Αποτελέσματα

Στις προηγούμενες υποενότητες, περιγράψαμε ορισμένα παραδείγματα εφαρμογής των προτεινόμενων μεθόδων σημασιολογικής κατάτμησης και αναδείξαμε τις λεπτομέρειες τους και τα προτερήματά τους σε σχέση με τον παραδοσιακό αλγόριθμο κατάτμησης RSST. Σε αυτή εδώ την παράγραφο θα παρουσιάσουμε τα αποτελέσματα για το σύνολο των εικόνων, την επίδραση του εννοιολογικού πλαισίου καθώς και την ευρωστία του συστήματος σε θορυβώδη είσοδο (λανθασμένη ταξινόμηση αρχικών περιοχών). Το σύνολο εκμάθησης των 120 εικόνων χρησιμοποιήθηκε για να υπολογιστούν οι βέλτιστες τιμές των παραμέτρων των επιμέρους αλγορίθμων.

Στον πίνακα 4.4 παρουσιάζουμε τα αποτελέσματα για το σύνολο των εικόνων που ανήκουν στο θεματικό πεδίο “παραλία”, τις επιδόσεις κάθε αλγορίθμου για κάθε έννοια όσο και συνολικά. Τα νούμερα του πίνακα είναι ο μέσος όρος των *scores* όλων των εικόνων. Οι έξι στήλες αντιστοιχούν στα αποτελέσματα έξι διαφορετικών αλγορίθμων: (α) κατάτμηση με μετασχηματισμό watershed και ταξινόμηση περιοχών με ταίριασμα οπτικών περιγραφέων, (β) σημασιολογικού πλημμυρισμού SW , (γ) σημασιολογικού πλημμυρισμού και χρήση εννοιολογικού πλαισίου $SW + C$, (δ) κατάτμηση με RSST και ταξινόμηση περιοχών με ταίριασμα οπτικών περιγραφέων, (ε) σημασιολογικό $SRSST$ και τέλος (στ) σημασιολογικό RSST και χρήση εννοιολογικού πλαισίου $SRSST + C$.

Όπως είναι φανερό, η έννοια ΟΥΡΑΝΟΣ απολαμβάνει την καλύτερη επίδοση στην αναγνώριση και τον εντοπισμό, καθώς οποιαδήποτε μεθοδολογία και να χρησιμοποιηθεί επιτυγχάνεται επίδοση μεγαλύτερη του 0.9. Αρκετά καλά είναι τα αποτελέσματα και για τις περιοχές που χαρακτηρίζονται από τις έννοιες ΘΑΛΑΣΣΑ και ΑΜΜΟΣ, καθώς πρόκειται για έννοιες που μπορούν να χαρακτηριστούν επιτυχώς από το χρώμα και την υφή, ενώ δεν έχουν πολύ μεγάλη ποικιλία στην εμφάνισή τους. Αυτό συμβαίνει έντονα στην περίπτωση των ανθρώπων, καθώς η ποικιλία στο σχήμα τους, στο



Σχήμα 4.17: Μελέτη επίδρασης θορύβου εισόδου στην τελική ταξινόμηση.

μέγεθός τους και το αν είναι ντυμένοι ή όχι επηρεάζει αρνητικά την επίδοση του συστήματός μας. Τέλος, στις περιπτώσεις των εννοιών ΒΡΑΧΟΣ και ΒΛΑΣΤΗΣΗ ισχύουν εν μέρει τα παραπάνω, γεγονός που αντανακλά στην ποιότητα των αποτελεσμάτων τους, η οποία ποσοτικά βρίσκεται πιο πάνω από τα αντίστοιχα νούμερα για την έννοια ΑΝΘΡΩΠΟΣ.

Μελετώντας τα αποτελέσματα στο σύνολο των εννοιών, παρατηρούμε ότι η χρήση σημασιολογικής πληροφορίας για την κατάτμηση μίας εικόνας βελτιώνει σε ποσοστό μεγαλύτερο του 10% τα αποτελέσματα του ορθού εντοπισμού και της αναγνώρισης σημασιολογικών οντοτήτων. Πιο συγκεκριμένα, συγκρίνοντας το μετασχηματισμό watershed με τον αλγόριθμο σημασιολογικού πλημμυρισμού *SW* έχουμε συνολική αύξηση 11.6% και 14.5% με ή χωρίς την επιπλέον χρήση του εννοιολογικού πλαισίου. Ανάλογα είναι τα νούμερα και στην περίπτωση του RSST και του *SRSST*, όπου η αύξηση είναι 10% και 11.4%, πάλι με ή χωρίς τη χρήση του εννοιολογικού πλαισίου.

Μελέτη Επίδρασης Θορύβου

Καθώς τόσο οι δύο αλγόριθμοι σημασιολογικής κατάτμησης όσο και αυτός του εννοιολογικού πλαισίου εξαρτώνται από την αρχικοποίηση του *ARG*, δηλαδή, την αρχική ταξινόμηση των περιοχών που προκύπτουν από την αρχική υπερκατάτμηση, οφείλουμε να ερευνήσουμε την απόκριση του συστήματός μας σε θορυβώδη είσοδο. Ως είσοδο θεωρούμε τα ασαφή σύνολα εννοιών της κάθε περιοχής και ως έξοδο την τελική μάσκα κατάτμησης και τα νέα ασαφή σύνολα εννοιών των νέων περιοχών. Για αυτό το σκοπό, προσθέσαμε γκαουσιανό θόρυβο πολλαπλών βαθμίδων (μεταβάλλοντας την διακύμανση σ της γκαουσιανής κατανομής) στις τιμές των βαθμών συμμετοχής των ασαφών συνόλων εισόδου, εκτελέσαμε, υπό ακριβώς τις ίδιες συνθήκες, τα πειράματά μας για το σύνολο των εικόνων και υπολογίσαμε το βαθμό επίδοσης *score* της κάθε εικόνας.

Η επιλογή του γκαουσιανού μοντέλου θορύβου δεν έγινε για κάποιο συγκεκριμένο λόγο, καθώς θεωρούμε ότι οποιοδήποτε μοντέλο τυχαίας μεταβολής (και όχι

κάποιας συγκεκριμένης κατανομής) της εισόδου θα μπορούσε να χρησιμοποιηθεί για να εξομοιώσει τη λανθασμένη αρχική κατάτμηση και ταξινόμηση. Για τον προσθετικό γκαουσιανό θόρυβο, δοκιμάσαμε διάφορες τιμές της διακύμανσης και πιο συγκεκριμένα $\sigma \in [0, 0.3]$ με βήμα 0.025. Για $\sigma = 0$ έχουμε απουσία θορύβου, ενώ ορίσαμε μέγιστη τιμή $\sigma = 0.3$ καθώς πειραματικά παρατηρήσαμε ότι για μεγαλύτερες τιμές οι θορυβώδεις τιμές συμμετοχής έπεφταν εκτός του πεδίου $[0, 1]$.

Το αποτέλεσμα του πειράματος αυτού, απεικονίζεται στο διάγραμμα του σχήματος 4.17. Πιο συγκεκριμένα, στο διάγραμμα φαίνονται οι τιμές *score* σε διάφορα επίπεδα θορύβου στην είσοδο, για κάθε έννοια ξεχωριστά καθώς και ο σταθμισμένος (με βάση τη συχνότητα εμφάνισής των) μέσος όρος αυτών (παχιά γραμμή), για τις 443 εικόνες του θεματικού πεδίου “παραλία”. Μπορούμε να παρατηρήσουμε ότι για χαμηλά έως μεσαία επίπεδα θορύβου ($\sigma < 0.15$) το σύστημα ανταποκρίνεται αρκετά καλά, επιδεικνύοντας μία σταθερή και εύρωστη συμπεριφορά.

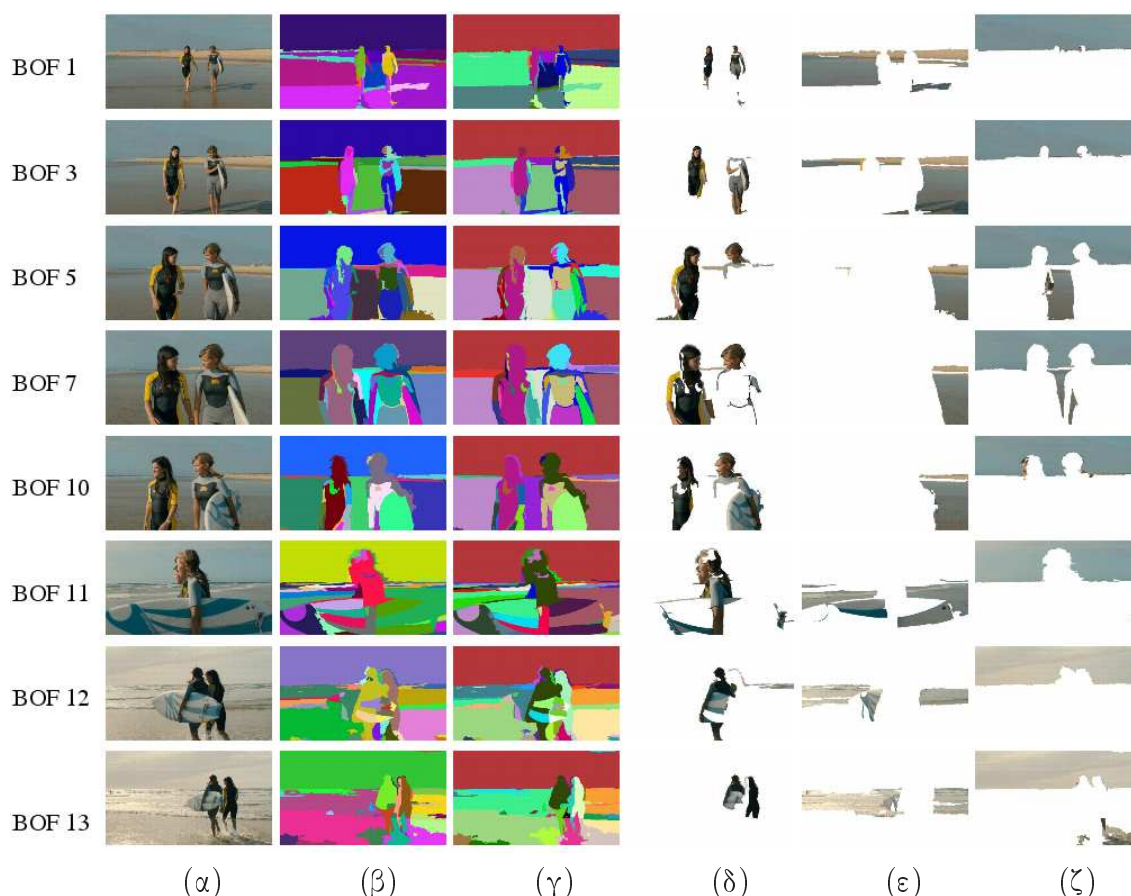
Εν προκειμένω, για μικρές τιμές θορύβου, η συνολική επίδοση του συστήματος παραμένει αναλλοίωτη γύρω στο 0.7 με 0.8. Οι επιδόσεις για κάθε έννοια, ανεξάρτητα, κυμαίνονται από 0.55 έως 0.95, τιμές που κρίνονται ικανοποιητικές δεδομένης της ποικιλίας των εικόνων. Πιο συγκεκριμένα οι έννοιες ΟΥΡΑΝΟΣ και ΘΑΛΑΣΣΑ αποδεικνύονται πολύ ανθεκτικές ακόμα και σε μεγάλα επίπεδα θορύβου. Όπως αναμέναμε, για μεγάλες τιμές θορύβου, π.χ. $\sigma > 0.2$, τα δεδομένα εισόδου είναι σχεδόν τυχαία, με αποτέλεσμα το προτεινόμενο σύστημα να οδηγείται σε μη-ντετερμινιστικά αποτελέσματα (τυχαία αυξομειώση των καμπυλών για $\sigma > 0.2$). Συμπερασματικά, ακόμα και όταν η αρχική ταξινόμηση αποτυγχάνει σε ένα βαθμό, το σύστημά μας εκμεταλλευόμενο τη σημασιολογική πληροφορία όλης της γειτονιάς μίας περιοχής, επιτυγχάνει ικανοποιητικά τελικά αποτελέσματα.

4.6 Πειραματικά Αποτελέσματα σε Ακολουθίες Βίντεο

Σε αυτή την ενότητα παρουσιάζουμε ενδεικτικά αποτελέσματα της προτεινόμενης μεθόδου σημασιολογικής κατάτμησης ακολουθιών βίντεο, τα οποία έχουν διάρκεια από 100 έως 650 καρέ. Το σημασιολογικό λεξιλόγιο περιλαμβάνει τις έννοιες ΘΑΛΑΣΣΑ, ΟΥΡΑΝΟΣ, ΒΛΑΣΤΗΣΗ και ΑΝΘΡΩΠΟΣ. Για τη διάρκεια του BoF δοκιμάσαμε διάφορες τιμές και καταλήξαμε ότι η βέλτιστη τιμή ήταν $|B| = 10$.

Η πρώτη ακολουθία βίντεο (σχήμα 4.18) ανήκει στη θεματική περιοχή “παραλία” και απεικονίζει δύο κοπέλες που περπατούν δίπλα στη θάλασσα. Αρχικά οι κοπέλες εμφανίζονται στο βάθος πλησιάζοντας την κάμερα (BoFs 1-5), ακολουθεί ένα διάστημα όπου φαίνονται σε κοντινή απόσταση (BoFs 6-10) και τέλος η κάμερα περιστρέφεται γρήγορα 180° και δείχνει τις κοπέλες από πίσω (BoFs 11-15). Οι έννοιες που εντοπίζονται στα επιλεγμένα καρέ είναι οι ΑΝΘΡΩΠΟΣ, ΘΑΛΑΣΣΑ και ΟΥΡΑΝΟΣ. Οι κύριοι όγκοι που εντοπίζονται είναι η κοπέλα στα αριστερά (έννοια ΑΝΘΡΩΠΟΣ), η δεξιά περιοχή της θάλασσας (έννοια ΘΑΛΑΣΣΑ) και η περιοχή του ουρανού (έννοια ΟΥΡΑΝΟΣ).

Η πρώτη παρατήρηση είναι ότι η περιοχή του ουρανού έχει αναγνωριστεί σε όλα τα BoFs και μάλιστα ως κύριος όγκος και επειδή επαληθεύει το κριτήριο (4.11)(α) οι διαδοχικοί όγκοι συγχωνεύονται και αποτελούν μία οντότητα καθ'όλη τη διάρκεια του πλάνου. Στην περίπτωση των όγκων που έχουν χαρακτηριστεί από την έννοια ΘΑΛΑΣΣΑ (δεξιά από τις κοπέλες) η συγχώνευση μεταξύ διαδοχικών BoFs βασίστηκε και πάλι στη σημασιολογική ομοιότητα του κριτηρίου (4.11)(α). Παρατηρούμε ότι στη στήλη (β) του σχήματος 4.18 οι όγκοι που αντιστοιχούν στο κομμάτι αυτό της θάλασ-

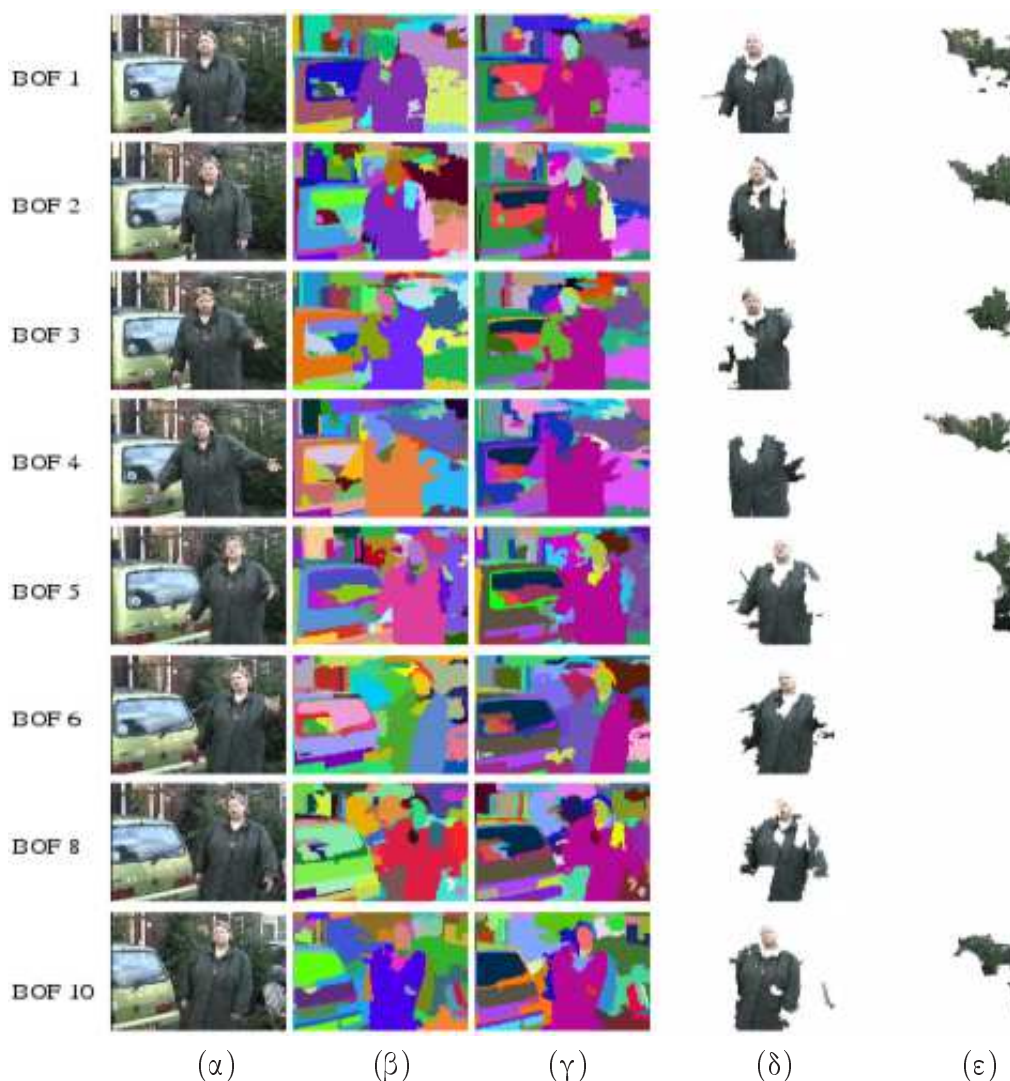


Σχήμα 4.18: (α) Καρέ σε διαφορετικά BoFs. (β) Αρχική χωροχρονική κατάτμηση κάθε BoF. (γ) Σημασιολογική κατάτμηση και συνένωση BoFs. Οι όγκοι συγχωνεύονται και επεκτείνονται σε όλη τη διάρκεια του πλάνου (όπως αποτυπώνεται από το σταθερό χρωματισμό). (δ) Εντοπισμός της έννοιας ΑΝΘΡΩΠΟΣ. (ε) Εντοπισμός της έννοιας ΘΑΛΑΣΣΑ. (ζ) Εντοπισμός της έννοιας ΟΥΡΑΝΟΣ.

σας έχουν εξαχθεί ορθά, αλλά δεν υπάρχει η χρονική συνέχεια σε όλο το πλάνο, κάτι που επιτυγχάνεται με το ταίριασμα των όγκων σε διαδοχικά BoFs όπως περιγράψαμε στην υποενότητα 4.3.4.1 και απεικονίζεται στη στήλη (γ) με τον κοινό χρωματισμό σε όλα τα BoFs. Ενδιαφέρον παρουσιάζει και η περίπτωση της περιοχής της θάλασσας αριστερά από τις κοπέλες. Όπως φαίνεται από τη στήλη (ε) του σχήματος 4.18, η αναγνώριση της έννοιας ΘΑΛΑΣΣΑ απέτυχε, αλλά το ταίριασμα των όγκων στα διαδοχικά BoFs επετεύχθη χάρη στην οπτική ομοιότητα και το κριτήριο (4.11)(γ).

Ο εντοπισμός και η αναγνώριση της έννοιας ΑΝΘΡΩΠΟΣ είναι πιο δύσκολη, καθώς η οπτική της περιγραφή είναι ιδιαίτερα περίπλοκη. Στο BoF 1 οι δύο σιλουέτες των κοριτσιών εντοπίζονται σωστά και εξάγεται ένας όγκος για την κάθε μία. Ο όγκος της αριστερά κοπέλας του BoF 1 συγχωνεύεται ορθώς με τους αντίστοιχους μέχρι το BoF 10, όταν και εξαφανίζεται λόγω επικάλυψης από την άλλη κοπέλα και από το BoF 12 μέχρι το τέλος επανεντοπίζεται. Καθώς δεν υποστηρίζεται το ταίριασμα όγκων μεταξύ μη διαδοχικών BoFs, σε τέτοιες περιπτώσεις (πρόβλημα επικάλυψης) δεν είναι δυνατή η ταυτοποίηση με το αντικείμενο που επανεμφανίζεται στη σκηνή.

Στο δεύτερο παράδειγμα (σχήμα 4.19) φαίνεται μία γυναίκα σε κοντινό πλάνο μπροστά από ένα αυτοκίνητο και ένα σπίτι. Οι έννοιες που εντοπίζονται είναι οι ΑΝΘΡΩΠΟΣ και ΒΛΑΣΤΗΣΗ, ενώ οι έννοιες ΑΥΤΟΚΙΝΗΤΟ και ΚΤΗΡΙΟ, αν και παρούσες, είναι τεμαχισμένες σε πολλούς όγκους και δεν ήταν δυνατό να αναγνωριστούν. Μπο-



Σχήμα 4.19: (α) Καρέ σε διαφορετικά BoFs. (β) Αρχική χωροχρονική κατάτμηση κάθε BoF. (γ) Σημασιολογική κατάτμηση και συνένωση BoFs. Οι όγκοι συγχωνεύονται και επεκτείνονται σε όλη τη διάρκεια του πλάνου (όπως αποτυπώνεται από το σταθερό χρωματισμό). (δ) Εντοπισμός της έννοιας ΑΝΘΡΩΠΟΣ. (ε) Εντοπισμός της έννοιας ΒΛΑΣΤΗΣΗ.

ρούμε να παρατηρήσουμε ότι η εικονιζόμενη γυναίκα έχει κατατμηθεί σε περισσότερα από ένα μέρη, με κυριότερα αυτά του κεφαλιού και του παλτού της. Επειδή ο βαθμός συμμετοχής της έννοιας ΑΝΘΡΩΠΟΣ, όπως προέκυψε από τη διαδικασία της αναγνώρισης, ήταν σχετικά μικρός, οι εν λόγω όγκοι δε θεωρήθηκαν κύριοι και επομένως δεν συγχωνεύθηκαν ούτε εντός του κάθε BoF ούτε μεταξύ τους στα διαδοχικά BoFs. Επιπλέον δυσκολία αποτέλεσε και η λανθασμένη αρχική χωροχρονική κατάτμηση στο BoF 4, όπου το κεφάλι έχει συνενωθεί με το παρασκήνιο. Τέλος η βλάστηση έχει εντοπιστεί και κατατμηθεί ορθά, με εξαίρεση τα BoFs 6-8 όπου, λόγω επικάλυψης, δεν κατέστη δυνατός ο εντοπισμός τους.

Η αξιολόγηση των αποτελεσμάτων της σημασιολογικής κατάτμησης των παραπάνω ακολουθιών έγινε με δύο διαφορετικά κριτήρια για κάθε έννοια, όπως φαίνεται και στον πίνακα 4.5. Το μέτρο ορθότητας *Acc* [43] μετράει την ακρίβεια της κατάτμησης για κάθε έννοια ενοποιώντας την ακρίβεια και την ανάκληση. Η μετρική *score* δίνεται από την εξίσωση 4.17 και αναλύεται στην ενότητα 4.5.1. Η πρώτη παρατήρηση που μπορεί να γίνει είναι ότι η έννοια ΟΥΡΑΝΟΣ αναγνωρίζεται και κατατέμεται

Πίνακας 4.5: Αξιολόγηση σημασιολογικής κατάτμησης βίντεο.

		ΒΛΑΣΤΗΣΗ	ΑΝΘΡΩΠΟΣ	ΘΑΛΑΣΣΑ	ΟΥΡΑΝΟΣ	Συνολικά
Παράδ.1	<i>Acc</i>	x	0.74	0.87	0.96	0.89
	<i>Score</i>	x	0.62	0.65	0.78	0.71
Παράδ.2	<i>Acc</i>	0.81	0.86	x	x	0.84
	<i>Score</i>	0.55	0.64	x	x	0.61

καλύτερα, καθώς έχει τη μεγαλύτερη τιμή και με τις δύο μετρικές. Για την έννοια ΒΛΑΣΤΗΣΗ αν και η μετρική *Acc* είναι υψηλή, το *score* είναι αρκετά χαμηλότερο, γεγονός που εξηγείται από την ανεπαρκή κατάτμηση του σε ορισμένα BoFs. Η έννοια ΘΑΛΑΣΣΑ έχει υψηλές τιμές *Acc* και *score*, δεδομένης της σταθερής οπτικής της περιγραφής (χρώματος και υφής). Τέλος, στην περίπτωση της έννοιας ΑΝΘΡΩΠΟΣ, τα αποτελέσματα ήταν κάπως καλύτερα στο δεύτερο βίντεο (σχήμα 4.19) καθώς δεν υπήρχαν τα προβλήματα επικάλυψης που παρουσιάστηκαν σε ορισμένα BoFs του πρώτου βίντεο.

4.7 Συμπεράσματα

Οι μέθοδοι που παρουσιάσαμε στο παρόν κεφάλαιο μπορούν να συνδράμουν στην υλοποίηση έξυπνων και αποδοτικών περιβαλλόντων ανάλυσης εικόνων. Δείξαμε ότι η κατάτμηση εικόνων και ο εντοπισμός και αναγνώριση εννοιών μπορούν να ενοποιηθούν ως πρόβλημα και να αντιμετωπιστούν παράλληλα, πετυχαίνοντας σημαντική βελτίωση σε σχέση με σειριακές προσεγγίσεις όπου η κατάτμηση προηγείται της αναγνώρισης. Η κύρια συνεισφορά μας είναι η προσέγγιση της ταυτόχρονης κατάτμησης και αναγνώρισης μίας εικόνας, η οποία παρέχει μία νέα οπτική στις τεχνικές εντοπισμού και αναγνώρισης αντικειμένων. Δύο παράγωγα αυτής της ιδέας είναι (α) οι δύο αλγόριθμοι σημασιολογικής επέκτασης περιοχών, ήτοι η σημασιολογικού πλημμυρισμού *SW* και ο σημασιολογικός *SRSSST*, οι οποίοι εξετάστηκαν σε ένα σύνολο πραγματικών εικόνων και (β) η χωροχρονική κατάτμηση βίντεο, με τη σημασιολογική συνένωση διαδοχικών μπλοκ από καρέ (BoFs).

Δευτερεύουσα συνεισφορά είναι το κριτήριο επίδοσης ενός συστήματος εντοπισμού εννοιών σε μία εικόνα. Προτείναμε ένα μέτρο το οποίο λαμβάνει υπόψιν του τόσο την ακριβή κατάτμηση της εικόνας σε σχέση με το σύνολο δεδομένων αλήθειας, όσο και την ορθή αναγνώριση της έννοιας που απεικονίζεται σε κάθε περιοχή. Πρόκειται για ένα μέτρο επιτυχούς ταξινόμησης βασισμένο σε περιοχές και όχι στο σύνολο της εικόνας, όπως συχνά γίνεται σε περιπτώσεις ταξινόμησης εικόνων. Πρόκειται για ένα πιο αυστηρό μέτρο, καθώς απαιτεί τον ακριβή εντοπισμό της έννοιας στην εικόνα και όχι απλά την ορθή απάντηση στο ερώτημα αν υπάρχει ή όχι η έννοια κάπου μέσα στην εικόνα.

□

Κεφάλαιο 5

Σημασιολογική Ερμηνεία και Συλλογιστική: Οπτική Λογική

5.1 Εισαγωγή

Το πολυμεσικό περιεχόμενο που παράγεται και διανέμεται τα τελευταία χρόνια είναι πραγματικά τεράστιο και οι τεχνολογίες αναζήτησης δεν έχουν καταφέρει να ικανοποιήσουν τις ανάγκες του τελικού χρήστη, για δύο κυρίως λόγους. Ο πρώτος είναι ότι η εργασία της χειρωνακτικής τεκμηρίωσης των πολυμέσων στα τωρινά πρότυπα μεταδεδομένων είναι δύσκολη και χρονοβόρα. Ο δεύτερος οφείλεται στο ότι είναι οι γλώσσες περιγραφής που χρησιμοποιούνται από τα πρότυπα μεταδεδομένων (όπως για παράδειγμα το πρότυπο MPEG-7 και το Dublin Core), βασίζονται στην XML(s), η οποία δεν παρέχει δυνατότητες σημασιολογικής αναπαράστασης [28, 55, 54]. Το ερώτημα λοιπόν είναι πως μπορεί κάποιος να δεικτοδοτήσει, να αναζητήσει και να ανακτήσει πολυμεσικό περιεχόμενο με βάση το σημασιολογικό του περιεχόμενο: ένα ερώτημα το οποίο με άλλα λόγια εκφράζεται ως η εξαγωγή και αναπαράσταση των σημαντικών εννοιολογικών περιεχομένων των δεδομένων, με τέτοιο τρόπο, ώστε να είναι εφικτή η κατανόηση των τόσο από τον τελικό χρήστη, όσο και από τον υπολογιστή [136, 87, 122].

Τα πολυμεσικά αρχεία είναι, εν γένει, σύνθετα χωρο-χρονικά σήματα που περιέχουν τόσο αριθμητική όσο και σημασιολογική πληροφορία. Για αυτόν το λόγο, πολλές μέθοδοι ανάλυσης και τεκμηρίωσης, που έχουν προταθεί στη βιβλιογραφία, υποβοηθούνται από κάποιου είδους γνώση [122, 56, 132]. Η ιδέα αυτής της προσέγγισης είναι καταρχάς να ερμηνευτεί το περιεχόμενο με βάση τα επιμέρους στοιχεία που το συνθέτουν (ήχος, βίντεο, φωνή, εικόνα, κείμενο) σε πολλαπλά επίπεδα λεπτομέρειας και έπειτα να καθοριστούν οι πηγές σημασιολογικής πληροφορίας (σκηές ενός βίντεο, πλάνα, καρέ, συγκεκριμένα αντικείμενα, πλαίσιο κειμένου, υπότιτλοι, αφήγηση, ηχητικά εφέ κ.ά.). Τελικά η σημασιολογική κατανόηση επιτυγχάνεται από αυτές τις πηγές πληροφορίας και τις αλληλεξαρτήσεις τους.

Για παράδειγμα ας υποθέσουμε, χωρίς απώλεια της γενικότητας, ότι θέλουμε να αναλύσουμε μία φωτογραφία που κατά κύριο λόγο απεικονίζει ένα αυτοκίνητο. Ένας αλγόριθμος κατάτμησης με βάση το χρώμα εξάγει μία περιοχή που απεικονίζει ένα κομμάτι του αυτοκινήτου, μία πηγή πληροφορίας a (χωρίς φυσικά να έχουμε ακόμα αναγνωρίσει ότι πρόκειται για αυτοκίνητο). Κατόπιν, εξάγονται ορισμένα προκαθορισμένα χαρακτηριστικά (οπτικά χαρακτηριστικά που περιλαμβάνουν πληροφορία για τη γεωμετρία της περιοχής, το χρώμα της, την υφή κ.ά.) και αποθηκεύονται σε κάποια

δομή αναπαράστασης χαρακτηριστικών για εικόνες (όπως, για παράδειγμα, ο σχεσιακός γράφος χαρακτηριστικών *ARG*). Το επόμενο βήμα είναι η αναγνώριση με βάση το βέλτιστο ταίριασμα των χαρακτηριστικών της περιοχής και των χαρακτηριστικών μοντέλων αντικειμένων που αποτελούν ένα είδος πρότερης γνώσης. Εάν επιτευχθεί το ταίριασμα με την ετικέτα *ΑΥΤΟΚΙΝΗΤΟ*, τότε δημιουργείται μία σημασιολογική σχέση μεταξύ της περιοχής *a* και της έννοιας *ΑΥΤΟΚΙΝΗΤΟ*. Συνεχίζοντας αυτή τη διαδικασία για πολλές πηγές πληροφορίας μπορούμε να συνδέσουμε επιπλέον ετικέτες οι οποίες προστίθενται στη συνολική περιγραφή της εικόνας.

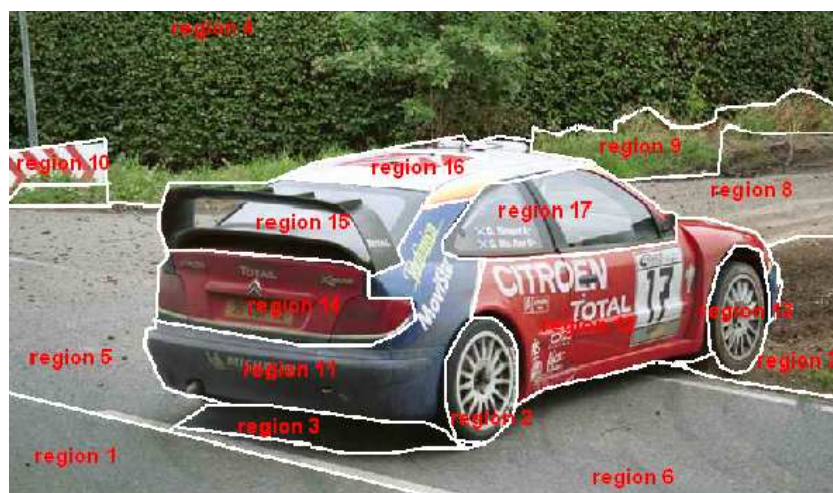
Αυτή η προσέγγιση, αν και μπορεί να εφαρμοστεί σε εικόνες με σχετικά μικρό αριθμό αντικειμένων, ή ισοδύναμα λίγες πηγές πληροφορίας, για περιπτώσεις όπου εμφανίζεται μεγάλος αριθμός αντικειμένων και εννοιών είναι πολύ αισιόδοξη ώστε να αναμένουμε ικανοποιητικά αποτελέσματα. Δύο σημαντικοί λόγοι για αυτό (μεταξύ άλλων φυσικά) είναι πρώτον ο εντοπισμός των σωστών πηγών πληροφορίας (που στο παραπάνω παράδειγμα βασίζεται στη διαδικασία της κατάτμησης και της εξαγωγής κατάλληλων περιοχών) και δεύτερον ο καθορισμός της σημασιολογικής ερμηνείας δια της διαδικασίας αναγνώρισης (όπως είναι το ταίριασμα με μοντέλα αντικειμένων).

Στο παρόν κεφάλαιο προτείνουμε μία ενιαία αντιμετώπιση των δύο παραπάνω προβλημάτων, χρησιμοποιώντας μία τυπική αναπαράσταση γνώσης και ένα πλαίσιο συλλογιστικής που καθορίζουν την περιγραφή και την οπτικοποίηση των υπό ανίχνευση εννοιών, εισάγοντας τον όρο της *οπτικής λογικής* (*visual logics*). Πιο συγκεκριμένα, στην ενότητα 5.2 θα διατυπωθεί το πρόβλημα εκτενέστερα και πιο τυπικά, καθώς και η δομή της προσέγγισης μας για την επίλυσή του. Στη συνέχεια στην ενότητα 5.3 θα περιγράψουμε το πλαίσιο της οπτικής λογικής μέσα από τη σκοπιά των ασαφών περιγραφικών λογικών. Έπειτα, στην ενότητα 5.4 περιγράφουμε τη διαδικασία εξαγωγής οπτικών χαρακτηριστικών και την αντιστοίχηση τους με έννοιες του πλαισίου της οπτικής λογικής, με χρήση ασαφών συνόλων και υπηρεσιών συλλογιστικής. Τέλος στην ενότητα 5.5 δίνονται κάποια ενδεικτικά αποτελέσματα, ενώ στην ενότητα 5.6 παρουσιάζονται παλαιότερες συναφείς εργασίες.

5.2 Προσδιορισμός του Προβλήματος

Στις παρακάτω παραγράφους θα διατυπώσουμε πιο τυπικά το πρόβλημα της σημασιολογικής κατάτμησης και κατανόησης εικόνων και θα παρουσιάσουμε την προσέγγισή μας και την αρχιτεκτονική του μοντέλου μας, ενώ οι ορισμοί και οι λεπτομέρειες περιγράφονται στις επόμενες ενότητες.

Ένα σύστημα ανάλυσης εικόνας ερμηνεύει μία αριθμητική (πιθανά και συντακτική) αναπαράσταση της εικόνας σε μία συμβολική (πιθανά και σημασιολογική) περιγραφή του περιεχομένου της. Στη δική μας προσέγγιση υιοθετούμε αναπαράσταση της εικόνας με βάση τις περιοχές, χρησιμοποιώντας το γράφο *ARG*, όπως αυτός ορίστηκε στην ενότητα 2.3.1, επιλογή η οποία παρέχει μία δομή για ένα σύνολο ισχυρισμών που επιδέχονται πολλαπλές διερμηνείες. Από την άλλη, για την συμβολική περιγραφή του περιεχομένου υιοθετήσαμε το φορμαλισμό των περιγραφικών λογικών (*description logics - DLs*), επιλογή, η οποία αποτελεί τη βάση της γλώσσας *OWL* που είναι η πρότυπη γλώσσα της *W3C* για την αναπαράσταση οντολογιών [12] και επομένως μπορεί να χρησιμοποιηθεί απευθείας ως περιγραφή περιεχομένου στο σημασιολογικό ιστό (*semantic web*) [123]. Προς αυτή την κατεύθυνση έχει γίνει μεγάλη προσπάθεια να μεταφερθούν και να ερμηνευτούν οι τυπικές μορφές μεταδεδομένων εικόνων (*MPEG-7, VRA, Dublin Core κ.ά.*) σε περιγραφικές λογικές [137]. Επιγραμματικά,



Σχήμα 5.1: Κατάτμηση εικόνας με βάση τη χρωματική πληροφορία.

να αναφέρουμε ορισμένα επιπλέον πλεονεκτήματα της χρήσης των περιγραφικών λογικών στη σημασιολογική αναπαράσταση. Πρώτον, οι βάσεις γνώσεις που βασίζονται σε περιγραφικές λογικές παρέχουν τη δυνατότητα για υπηρεσίες συλλογιστικής, οι οποίες μπορούν να χρησιμοποιηθούν για τον εμπλουτισμό της συμβολικής περιγραφής. Δεύτερον, έχουν καθοριστεί επεκτάσεις στις περιγραφικές λογικές που καλύπτουν τη διαχείριση αβεβαιότητας [124], η οποία ενυπάρχει ως χαρακτηριστικό σε οποιοδήποτε σύστημα ανάλυσης εικόνας. Τέλος, οι περιγραφικές λογικές βασίζονται στην υπόθεση ανοικτού κόσμου (open-world assumption), μία πραγματικά σημαντική ιδιότητα της λογικής πρώτης τάξης (first order logic) και των σημασιολογικών της τύπων.

Ας προσπαθήσουμε να αναλύσουμε το πρόβλημα. Στο σχήμα 5.1 μπορούμε να δούμε μία εικόνα και την κατάτμησή της σε έναν αριθμό περιοχών. Κάθε περιοχή αναπαριστά μία ομάδα εικονοστοιχείων τα οποία έχουν ένα ή περισσότερα κοινά χαρακτηριστικά. Συνήθως, αυτά τα χαρακτηριστικά είναι οπτικής φύσης, όπως το χρώμα (για παράδειγμα το γκρι χρώμα της περιοχής s_1 της εικόνας του σχήματος 5.1). Ένα χαρακτηριστικό, που μοιράζονται από κοινού εικονοστοιχεία μιας ή περισσότερων περιοχών, μπορεί να εκφράζει τη συμμετοχή στην οπτικοποίηση μίας έννοιας. Για παράδειγμα, η περιοχή s_2 δεν χαρακτηρίζεται από κάποια ομοιομορφία σε κάποιο αριθμητικό χαρακτηριστικό (όπως ήταν το χρώμα για την s_1), αλλά μοιράζονται το χαρακτηριστικό της απεικόνισης της έννοιας ΡΟΔΑ, έχουν δηλαδή ένα κοινό σημασιολογικό χαρακτηριστικό. Επιπρόσθετα, η συγκεκριμένη περιοχή δεν έχει μόνο την έννοια ΡΟΔΑ ως κοινό σημασιολογικό χαρακτηριστικό, αλλά συμμετέχει εξίσου μαζί με άλλες περιοχές στην οπτικοποίηση των εννοιών ΑΥΤΟΚΙΝΗΤΟ, ΡΑΛΙ κ.ά. Χρησιμοποιώντας αυτή τη λογική μπορούμε να δούμε κάθε περιοχή ως πηγή πληροφορίας, από την οποία μπορούμε να εξάγουμε οπτική πληροφορία από τα εικονοστοιχεία που τη συνθέτουν (όπως είναι π.χ. οι οπτικοί περιγραφείς), αλλά ορισμένες φορές μπορούμε επίσης να εξάγουμε και σημασιολογική πληροφορία (όπως είναι π.χ. η έννοια ή οι έννοιες τις οποίες οπτικοποιεί). Το πρώτο επιτυγχάνεται μέσω διαδικασίας εξαγωγής χαρακτηριστικών, ενώ το δεύτερο συνήθως μέσω διαδικασίας αναγνώρισης. Είναι προφανές ότι αν και η πρώτη διαδικασία μπορεί πάντα να οδηγήσει με επιτυχία στην οπτική περιγραφή της περιοχής, η επιτυχία της αναγνώρισης βασίζεται κυρίως στη σημασιολογική ομοιογένεια της περιοχής. Για παράδειγμα, ο οπτικός περιγραφέας κύριων χρωμάτων της περιοχής s_{15} της εικόνας του σχήματος 5.1 μπορεί εύκολα να

εξαχθεί, αλλά δεν είναι καθόλου σαφές ποια είναι η έννοια (αν υπάρχει) την οποία οπτικοποιούν από κοινού, γεγονός που οδηγεί σε πιθανή αποτυχία του αλγορίθμου αναγνώρισης.

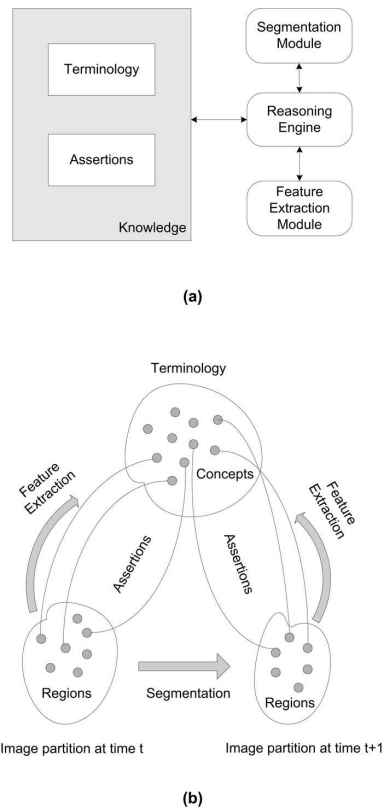
Σύμφωνα με τον παραπάνω φορμαλισμό, η διαδικασία της κατάτμησης μπορεί να οριστεί ως ο καθορισμός των κατάλληλων πηγών πληροφορίας, ενώ η διαδικασία της αναγνώρισης ως η εξαγωγή της σημασιολογικής πληροφορίας από τις πηγές αυτές. Αυτή η ιεραρχική προσέγγιση είναι προβληματική σε περιπτώσεις όπου ο αριθμός των αντικειμένων σε μία εικόνα είναι μεγάλος, καθώς τότε είναι απαραίτητο να συμπεριληφθούν τεχνικές αναγνώρισης στη διαδικασία της κατάτμησης ούτως ώστε να εξαχθούν σημασιολογικά ορθές περιοχές. Κατ' αυτό τον τρόπο οι δυσκολίες και τα προβλήματα της αναγνώρισης μετατίθενται στην κατάτμηση. Στην προτεινόμενη τεχνική, προσπαθούμε να το αποφύγουμε υποστηρίζοντας την κατάτμηση με σημασιολογική ανάλυση βασισμένη σε υπηρεσίας συλλογιστικής στη σχετική βάση γνώσης.

Το σχήμα 5.2 συνοψίζει την προτεινόμενη ιδέα. Η γνώση του συστήματος αποτελείται από τρία μέρη: την ορολογία (terminology), τους ισχυρισμούς (assertions) και την μηχανή συλλογιστικής (reasoning engine). Η ορολογία είναι ένα σύνολο αξιωμάτων που παρέχουν ένα σύνολο οπτικών ορισμών και υπαγωγών (subsumptions) που περιλαμβάνουν τις έννοιες και τις σχέσεις τους στη βάση γνώσης. Ένα παράδειγμα τέτοιου αξιώματος θα μπορούσε να δηλώνει ότι η έννοια ΚΟΚΚΙΝΟΑΥΤΟΚΙΝΗΤΟ είναι ένα αυτοκίνητο χρώματος κόκκινου. Είναι σημαντικό να τονίσουμε ότι η ορολογία καθορίζει την αφηρημένη γνώση, η οποία όμως μπορεί να συνδεθεί με τον πραγματικό κόσμο· στην περίπτωση που εξετάζουμε, οι περιοχές της εικόνας αποτελούν αντικείμενα του κόσμου, αν και δεν είναι απαραίτητο να περιορίσουμε τον κόσμο μόνο σε ότι απεικονίζεται στην εικόνα.

Οι ισχυρισμοί συνδέουν την ορολογία με τον κόσμο (την εικόνα), συνδέουν δηλαδή τις έννοιες με τις πηγές πληροφορίας. Για παράδειγμα, ένας ισχυρισμός μπορεί να δηλώνει ότι η περιοχή s_1 πιθανώς είναι 'κόκκινου χρώματος' ή ότι οι περιοχές s_1 και s_2 είναι 'γειτονικές'. Στην ενότητα 5.3 θα δοθούν περισσότερες λεπτομέρειες στο πως τυποποιούνται αυτοί οι ισχυρισμοί.

Η μηχανή συλλογιστικής (σχήμα 5.2) είναι ένα σύστημα που, χρησιμοποιώντας την ορολογία και ένα σύνολο ισχυρισμών, μπορεί να απαντά σε ερωτήματα όπως "είναι η περιοχή s_1 αυτοκίνητο;". Ένα σημαντικό χαρακτηριστικό της μηχανής συλλογιστικής είναι η υπόθεση ανοικτού κόσμου. Σύμφωνα με αυτήν, ένα σύνολο περιοχών το ερμηνεύουμε ως ένα πιθανώς άπειρο σύνολο το οποίο δεν περιορίζεται από το σύνολο των ορατών περιοχών. Αυτό υπονοεί ότι στο ερώτημα 'υπάρχει κάποια περιοχή κοντά στην s_1 ;', η μηχανή θα απαντήσει "ΟΧΙ" αν δεν υπάρχει στη συγκεκριμένη κατάτμηση της εικόνας, ενώ θα απαντήσει "ΝΑΙ" αν η ύπαρξη μίας περιοχής s (η οποία δεν περιλαμβάνεται στην υπό εξέταση κατάτμηση) που να είναι 'κοντά' στην s_1 δεν προκαλεί ασυνέπεια στη βάση γνώσης. Το συγκεκριμένο χαρακτηριστικό είναι πολύ χρήσιμο σε ορισμένες περιπτώσεις, όπως για παράδειγμα στην περίπτωση της επικάλυψης (occlusion) όταν ένα αντικείμενο καλύπτεται μερικώς ή ολικώς από κάποιο άλλο καθώς βρίσκεται πίσω του στην 3Δ σκηνή.

Χρησιμοποιώντας την παραπάνω υποδομή γνώσης, η προτεινόμενη διαδικασία κατάτμησης βελτιώνει την αρχική διαμέριση της εικόνας, ακολουθώντας μία πρακτική συγχώνευσης περιοχών και αναπροσαρμογής του συνόλου των ισχυρισμών που διέπουν τη νέα διαμέριση. Χοντρικά, η διαδικασία ξεκινάει με την εξαγωγή ορισμένων χαρακτηριστικών χαμηλού επιπέδου της εικόνας και αναθέτει ένα σύνολο οπτικών περιγραφών για κάθε περιοχή της εικόνας. Στη συνέχεια, αυτά τα χαρακτηριστικά



Σχήμα 5.2: Αρχιτεκτονική του προτεινόμενου συστήματος οπτικής λογικής.

καθορίζουν και αρχικοποιούν το σύνολο των ισχυρισμών (π.χ. η περιοχή s_1 είναι ‘κόκκινου χρώματος’). Τέλος, με χρήση υπηρεσιών συλλογιστικής, υπολογίζεται η απόσταση από όλες τις γειτονικές περιοχές με μία μετρική που βασίζεται τόσο σε αριθμητικούς όσο και σε σημασιολογικούς όρους. Πρέπει να σημειωθεί ότι δεν είναι τόσο καθοριστικός παράγοντας η αρχική κατάτμηση της εικόνας (αν και θα επιτάχυνε τη σύγκλιση προς την τελική κατάτμηση), ούτε κάποιο κατώφλι για το κριτήριο τερματισμού· στην ακραία περίπτωση, μπορεί να θεωρηθεί ως αρχική διαμέριση της εικόνας όλο το σύνολο των εικονοστοιχείων και ως τελική διαμέριση όλη η εικόνα ως μία περιοχή.

5.3 Αξιωματικός Σκελετός και Ερμηνεία της Οπτικής Λογικής

5.3.1 Μοντέλο Γνώσης Οπτικής Πληροφορίας

Το πρώτο βήμα στον καθορισμός ενός κατάλληλου μοντέλου γνώσης είναι να προσδιορίσουμε τις σημασιολογικές οντότητες, τις οποίες ονομάζουμε *έννοιες*, που έχουν σημαντικό ρόλο στην κατανόηση του περιεχομένου μίας εικόνας. Όπως αναφέρθηκε προηγουμένως, η εικόνα θεωρείται μία συλλογή περιοχών, από τις οποίες είναι σχετικά απλό να εξαχθούν οπτικά μετρήσιμα χαρακτηριστικά· πολύ δυσκολότερο είναι

το έργο της εξαγωγής σημασιολογικής πληροφορίας και της αντιστοίχισης των περιοχών με έννοιες. Οι έννοιες αποτελούν τα μοναδιαία κατηγορήματα (unary predicates) και με κατάλληλο συνδυασμό τους μπορούμε να εξάγουμε πιο σύνθετες έννοιες και περιγραφές. Επιπρόσθετα, περιοχές των οποίων τα περιγράμματα άπτονται, θεωρούνται γειτονικές, χαρακτηριστικό που μπορεί να αναπαρασταθεί ως μία χωρική σχέση. Καθώς αυτού του είδους οι σχέσεις, ενώνουν δύο οντότητες, αναφέρονται ως ρόλοι. Έτσι δημιουργείται το τυπικό πλαίσιο της περιγραφής μίας εικόνας ως ένα σύνολο εννοιών \mathcal{C} και ρόλων \mathcal{R} .

Υπάρχουν έννοιες οι οποίες μπορούν να εξαχθούν απευθείας από εργαλεία ανάλυσης εικόνας, ενώ άλλες εξάγονται μέσω αξιωμάτων που συνδυάζουν έννοιες και ρόλους και ανήκουν σε ένα υψηλότερο επίπεδο εκφραστικότητας. Στη χαμηλότερη βαθμίδα αυτής της ιεραρχικής οπτικής λογικής, συγκαταλέγονται οι έννοιες τις οποίες ονομάζουμε διεγέρτες.

Ορισμός 5.1 (Διεγέρτης) Μία έννοια-διεγέρτης $\tau \in \mathcal{T}_C$ είναι η έννοια που μπορεί να αρχικοποιηθεί από μία διαδικασία εξαγωγής και ταιριάσματος χαρακτηριστικών.

Οι έννοιες-διεγέρτες είναι έννοιες που ανιχνεύονται με σχετικά απλές πράξεις πάνω στην εικόνα και τις περιοχές της, και σκοπός τους είναι να ξεκινήσουν, ή αλλιώς να διεγείρουν, μία διαδικασία επέκτασης της γνώσης βασισμένη στην οπτική λογική. Διαισθητικά, οι έννοιες-διεγέρτες (ή απλά διεγέρτες) αποτελούν τη διαχωριστική γραμμή μεταξύ της αναπαράστασης μίας περιοχής με αριθμητικά χαρακτηριστικά και της προσέγγισής της οπτικής λογικής. Επομένως, ο παραπάνω ορισμός έχει πολύ σημαντικό ρόλο, καθώς αποτελεί το πρώτο βήμα για την μετάβαση από τις αριθμητικές τιμές σε γλωσσικές τιμές, οι οποίες έχουν διαισθητικό νόημα και περιεχόμενο για έναν άνθρωπο. Οι έννοιες διεγέρτες ορίζονται αν θεωρήσουμε οπτικά χαρακτηριστικά όπως το χρώμα, η υφή ή το σχήμα μιας περιοχής. Έτσι, παραδείγματα διεγερτών σε σχέση με το χρώμα είναι οι ΚΟΚΚΙΝΟ, ΜΠΛΕ, ΑΝΟΙΧΤΟΜΠΛΕ, κλπ.· για την υφή ΤΡΑΧΥ, ΛΕΙΟ, ΚΑΤΕΥΘΥΝΤΙΚΟ, κλπ.· και τέλος για το σχήμα οι ΚΥΚΛΙΚΟ, ΕΠΙΜΗΚΕΣ, κλπ. Όλοι οι διεγέρτες, μαζί με τις μεθόδους εξαγωγής τους, παρουσιάζονται αναλυτικά στην παράγραφο 5.4. Παρόμοια με τους διεγέρτες, ορίζονται και οι ρόλοι-διεγέρτες \mathcal{T}_R : ένας ρόλος-διεγέρτης είναι ο ρόλος που μπορεί να αρχικοποιηθεί από μία διαδικασία εξαγωγής μετρήσιμων χαρακτηριστικών.

Συνεπώς, με βάση τους διεγέρτες και τους δυνατούς συνδυασμούς τους, είναι δυνατόν να οριστούν τα οπτικά αξιώματα (visual axioms), με τα οποία μπορούμε να ορίσουμε πιο εκφραστικές έννοιες. Ο πρώτος τύπος οπτικού αξιώματος που ορίζουμε, ονομάζεται απλό αξίωμα.

Ορισμός 5.2 (Απλό Αξίωμα) Ένα απλό αξίωμα είναι μία έκφραση \mathcal{E} που χρησιμοποιεί μόνο έννοιες-διεγέρτες: $C \equiv \mathcal{E}(\tau_1, \dots, \tau_n)$, όπου $\tau_i \in \mathcal{T}_C, 1 \leq i \leq n$ και C είναι μία έννοια τέτοια ώστε $C \neq \tau_i$.

Τα απλά αξιώματα συνιστούν το πρώτο βήμα για τον προσδιορισμό πιο σύνθετων και εκφραστικών εννοιών. Σημειώνεται ότι τα απλά αξιώματα δεν περιλαμβάνουν ρόλους, αλλά μόνο εκφράσεις μεταξύ διεγερτών. Αυτό σημαίνει ότι μία περιοχή χαρακτηρίζεται αποκλειστικά από τα δικές της ιδιότητες και δεν επηρεάζεται από αυτές των γειτονικών της περιοχών. Για παράδειγμα, η έκφραση $\text{ΒΟΥΝΟ} \equiv \mathcal{E}(\text{ΚΑΦΕ}, \text{ΤΡΑΧΥ})$ αποτελεί ένα απλό αξίωμα. Παρόμοια μπορούν να οριστούν και άλλες έννοιες με χρήση απλών αξιωμάτων, όπως ΚΑΘΑΡΟΣΟΥΡΑΝΟΣ και ΒΡΑΧΟΣ.

Τέλος, μπορούμε να ενσωματώσουμε και τους ρόλους στην κατασκευή των οπτικών αξιωμάτων. Οι ρόλοι χρησιμοποιούνται για να συνδέσουν δύο περιοχές και τις ιδιότητές τους και τις έννοιες-διεγέρτες που τις συνοδεύουν, για να δημιουργήσουν πιο σύνθετες εκφράσεις. Αυτά τα αξιώματα τα ονομάζουμε *δομικά αξιώματα*.

Ορισμός 5.3 (Δομικό Αξίωμα) Δομικό αξίωμα είναι μία έκφραση διεγερτών και ρόλων: $C \equiv \mathcal{E}(\tau_1, \dots, \tau_n, R_1, \dots, R_m)$, όπου $\tau_i \in \mathcal{T}_C, 1 \leq i \leq n$, $R_i \in \mathcal{T}_R, 1 \leq i \leq m$ και $C \notin \mathcal{T}_C$ είναι μία έννοια που δεν ανήκει στο σύνολο των διεγερτών.

Τα δομικά αξιώματα, καθώς περιλαμβάνουν ρόλους, μπορούν να διατυπώσουν σύνθετες έννοιες οι οποίες ορίζονται σε σχέση όχι μόνο με μία, αλλά με ένα συνδυασμό γειτονικών ή μη περιοχών. Κάθε μία από αυτές τις περιοχές πιθανόν να περιγράφεται από διαφορετικά οπτικά χαρακτηριστικά, όπως το χρώμα ή η υφή, ενώ επίσης μπορεί να οπτικοποιεί και διαφορετικές έννοιες· όλες αυτές όμως συμμετέχουν στην από κοινού οπτικοποίηση μίας τουλάχιστον έννοιας, πιθανότατα αποτελώντας μέρος ενός πιο σύνθετου αντικειμένου. Παραδείγματος χάριν, ο ορισμός της έννοιας ΔΕΝΤΡΟ είναι ένα δομικό αξίωμα βασισμένο στις έννοιες ΚΑΦΕ, ΦΥΛΛΑ και στο ρόλο ΓΕΙΤΝΙΑΖΕΙ: $\Delta\text{ΕΝΤΡΟ} \equiv \mathcal{E}(\text{ΚΑΦΕ}, \text{ΦΥΛΛΑ}, \text{ΓΕΙΤΝΙΑΖΕΙ})$. Το αξίωμα αυτό λέει ότι η έννοια ΔΕΝΤΡΟ ορίζεται ως μία περιοχή καφέ χρώματος (η οποία αντιστοιχεί στον κορμό του δέντρου) και από μία γειτονική περιοχή που χαρακτηρίζεται από την έννοια ΦΥΛΛΑ. Στον ορισμό της έννοιας ΔΕΝΤΡΟ θα μπορούσαμε να γίνουμε πιο αναλυτικοί, απαιτώντας συγκεκριμένου τύπου γειτνίαση (π.χ. χωρική όπως είναι η ΕΙΝΑΙΠΑΝΩ, ΕΙΝΑΙΑΡΙΣΤΕΡΑ, κλπ.) ή συμμετοχή περισσότερων περιοχών (π.χ. με χρήση της έννοιας ΚΛΑΔΙ). Είναι φανερό ότι οι περιοχές που μπορούν να συμμετέχουν στον ορισμό μίας έννοιας μέσω ενός δομικού αξιώματος μπορούν να έχουν εντελώς διαφορετικά οπτικά χαρακτηριστικά. Τα δομικά αξιώματα παρέχουν τη δυνατότητα να ορίσουμε πιο εκφραστικές έννοιες, με τη χρήση ρόλων όπως οι ΕΧΕΙΜΕΡΟΣ, ΕΧΕΙΑΜΕΣΟΜΕΡΟΣ, ΠΕΡΙΕΧΕΙ, ΕΙΝΑΙΠΑΝΩ κ.ά..

Συμπερασματικά, η παραπάνω αρχιτεκτονική, την οποία ονομάσαμε οπτική λογική, χρησιμοποιεί μία ιεραρχική αναπαράσταση γνώσης, ξεκινώντας από μετρήσιμα οπτικά χαρακτηριστικά (διεγέρτες) και φθάνοντας σε πιο σύνθετες και εκφραστικές έννοιες μέσω των οπτικών αξιωμάτων (απλών και δομικών). Η χρήση τυπικής γλώσσας αναπαράστασης γνώσης και τυπικής σημασιολογίας παρέχει δύο χρήσιμα πλεονεκτήματα: πρώτον, η βάση γνώσης που δημιουργείται μπορεί να μεταφερθεί, να επαναχρησιμοποιηθεί και να ερμηνευτεί εύκολα και ρητά σε ένα νέο σύστημα, και δεύτερον είναι δυνατόν να χρησιμοποιηθούν υπηρεσίες συλλογιστικής πάνω στη δεδομένη γνώση και η εξαγωγή συμπερασμάτων που οδηγεί στην εξαγωγή νέας πληροφορίας, διαδικασία που ονομάζουμε *οπτική συλλογιστική* (visual reasoning).

5.3.2 Περιγραφικές Λογικές για Αναπαράσταση Οπτικής Πληροφορίας

Οι περιγραφικές λογικές - ΠΛ (description logics - DLs) [11], αποτελούν την κατάλληλη γλώσσα αναπαράστασης γνώσης για να κωδικοποιηθεί η οπτική λογική [80, 90]. Οι ΠΛ είναι μία οικογένεια φορμαλισμού αναπαράστασης γνώσης βασισμένες στη λογική, σχεδιασμένες να αναπαριστούν και να συμπεραίνουν γνώση γύρω από ένα συγκεκριμένο πεδίο με δομημένο και σαφή τρόπο. Παρόμοια με το γνωσιακό μοντέλου του ανθρώπου, τα θεμελιώδη στοιχεία των ΠΛ είναι οι έννοιες και οι ρόλοι, τα οποία χρησιμοποιούνται σε αξιώματα και σχέσεις στιγμιοτύπων βάσεων γνώσεων.

Είναι προφανής η αντιστοιχία και η ομοιότητα με τη δομή του μοντέλου γνώσης της οπτικής λογικής που προτείναμε στην προηγούμενη ενότητα.

Ανάλογα με την εκφραστική δυνατότητα μίας γλώσσας, οι ΠΛ παρέχουν ένα σύνολο κατασκευαστών (constructors) για την κατασκευή σύνθετων εννοιών και ρόλων από απλούς. Πρώτα ορίζουμε το σύνολο των ατομικών εννοιών \mathcal{C} , των ατομικών ρόλων \mathcal{R} και των ατόμων \mathcal{I} . Στη συνέχεια μπορούμε να ορίσουμε σύνθετες έννοιες μέσω σύνθετων εκφράσεων. Αν υποθέσουμε ότι χρησιμοποιούμε τη γλώσσα *SHLN* DL [51]¹, τότε μία *SHLN*-έννοια ορίζεται επαγωγικά ως:

1. Αν $C \in \mathcal{C}$, τότε C είναι μία *SHLN*-έννοια,
2. Αν C και D είναι έννοιες, R είναι ρόλος και $n \in \mathbb{N}$, τότε $(\neg C)$, $(C \sqcup D)$, $(C \sqcap D)$, $(\forall R.C)$, $(\exists R.C)$, $(\geq nR)$ και $(\leq nR)$ είναι επίσης *SHLN*-έννοιες.

Επιπρόσθετα, αν R είναι ένας ρόλος, τότε και ο R^- είναι επίσης ρόλος και ονομάζεται ο αντίστροφος του R . Όπως μπορούμε να διαπιστώσουμε, η γλώσσα *SHLN* δεν προσφέρει τα μέσα για να δημιουργήσουμε σύνθετους ρόλους, δηλαδή υπάρχουν μόνο ατομικοί ρόλοι. Βεβαίως, πιο εκφραστικές ΠΛ, όπως αυτή που παρουσιάστηκε στην εργασία [80], δίνουν τη δυνατότητα χρήσης σύνθετων ρόλων. Επίσης, τα αξιώματα εννοιών είναι της μορφής $C \equiv D$ ή $C \sqsubseteq D$, όπου C, D είναι έννοιες και ερμηνεύονται ότι η έννοια C είναι ισοδύναμη ή υπό-έννοια της έννοιας D αντιστοίχως. Ένα σύνολο τέτοιων αξιωμάτων αποτελεί το *σώμα ορολογίας* (*TBox*) \mathcal{T} . Επιπλέον, μπορούμε να έχουμε αξιώματα ρόλων της μορφής $\text{Trans}(R)$ λέγοντας ότι ο ρόλος R είναι μεταβατικός και της μορφής $R \sqsubseteq S$ λέγοντας ότι ο ρόλος R είναι υπό-ρόλος του S . Το σύνολο των αξιωμάτων ρόλων αποτελεί το *σώμα ρόλων* (*RBox*) \mathcal{R} . Συνδυάζοντας τα αξιώματα εννοιών και ρόλων, δημιουργούμε την υπονοούμενη γνώση (intentional knowledge) ενός θεματικού πεδίου, η οποία μπορεί να χρησιμοποιηθεί για τη σημασιολογική ερμηνεία μίας εικόνας και των περιοχών που την αποτελούν.

Με την αντιστοίχιση της οπτικής λογικής και της σημειολογίας των ΠΛ που αναλύσαμε παραπάνω, μπορούμε να κάνουμε ορισμένες χρήσιμες παρατηρήσεις. Καταρχήν, οι διεγέρτες και οι ρόλοι-διεγέρτες παίζουν τον ίδιο ρόλο με τις ατομικές έννοιες και τους ατομικούς ρόλους των ΠΛ. Αυτό τονίζει ξανά ότι οι διεγέρτες αποτελούν το θεμελιακό στοιχείο της οπτικής λογικής και δεν μπορούν να οριστούν σε σχέση με άλλες έννοιες μέσω αξιωμάτων. Ακόμη, τα απλά αξιώματα είναι στην ουσία αξιώματα περιγραφικών λογικών που στη δεξιά μεριά περιλαμβάνουν έννοιες της μορφής $(\neg C)$, $(C \sqcup D)$, $(C \sqcap D)$, όπου C, D είναι διεγέρτες. Στο παράδειγμα της έννοιας *ΒΟΥΝΟ* το αξίωμα έχει ως εξής: $\text{ΒΟΥΝΟ} \equiv (\text{ΠΡΑΣΙΝΟ} \sqcup \text{ΚΑΦΕ}) \sqcap \text{ΤΡΑΧΥ}$. Από την άλλη, τα δομικά αξιώματα είναι αξιώματα ΠΛ που στη δεξιά μεριά περιλαμβάνουν τουλάχιστον μία έννοια της μορφής $(\forall R.C)$, $(\exists R.C)$, $(\geq nR)$ ή $(\leq nR)$, δηλαδή περιλαμβάνουν ρόλους που συνδυάζουν πολλές περιοχές για τον προσδιορισμό μίας σύνθετης έννοιας. Ένα τέτοιο παράδειγμα είναι η έννοια *ΔΕΝΔΡΟ* η οποία ορίζεται ως: $\text{ΔΕΝΔΡΟ} \equiv \text{ΚΑΦΕ} \sqcap \text{ΞΓΕΙΤΝΙΑΖΕΙ.ΦΥΛΛΑ}$. Τέλος, χρησιμοποιώντας τη σημειολογία των ΠΛ, η απαίτηση ότι μία έννοια C δεν ανήκει στο σύνολο των διεγερτών \mathcal{T}_C (ορισμοί 5.2 και 5.3) υποδηλώνει ότι το *TBox* είναι ακυκλικό, δηλαδή καμία έννοια δεν ορίζεται άμεσα ή έμμεσα μέσω του εαυτού της.

¹Σημειώνεται ότι μπορούν να χρησιμοποιηθούν και άλλα συστήματα περιγραφικών λογικών περισσότερο ή λιγότερο εκφραστικά.

Πίνακας 5.1: Σημασιολογία εννοιών και ρόλων.

Κατασκευαστής	Σύνταξη	Σημασιολογία
καθολική έννοια	\top	$\top^{\mathcal{I}}(a) = 1$
κενή έννοια	\perp	$\perp^{\mathcal{I}}(a) = 0$
γενικευμένη άρνηση	$\neg C$	$(\neg C)^{\mathcal{I}}(a) = c(C^{\mathcal{I}}(a))$
τομή (σύζευξη)	$C \sqcap D$	$(C \sqcap D)^{\mathcal{I}}(a) = t(C^{\mathcal{I}}(a), D^{\mathcal{I}}(a))$
ένωση (διάζευξη)	$C \sqcup D$	$(C \sqcup D)^{\mathcal{I}}(a) = u(C^{\mathcal{I}}(a), D^{\mathcal{I}}(a))$
υπαρξιακός περιορισμός	$\exists R.C$	$(\exists R.C)^{\mathcal{I}}(a) = \sup_{b \in \Delta^{\mathcal{I}}} \{t(R^{\mathcal{I}}(a, b), C^{\mathcal{I}}(b))\}$
περιορισμός τιμής	$\forall R.C$	$(\forall R.C)^{\mathcal{I}}(a) = \inf_{b \in \Delta^{\mathcal{I}}} \{\mathcal{J}(R^{\mathcal{I}}(a, b), C^{\mathcal{I}}(b))\}$
περιορισμός το-πολύ	$\leq pR$	$\inf_{b_1, \dots, b_{p+1} \in \Delta^{\mathcal{I}}} \mathcal{J}(t_{i=1}^{p+1} R^{\mathcal{I}}(a, b_i), u_{i < j} \{b_i = b_j\})$
περιορισμός το-λιγότερο	$\geq pR$	$\sup_{b_1, \dots, b_p \in \Delta^{\mathcal{I}}} t(t_{i=1}^p R^{\mathcal{I}}(a, b_i), t_{i < j} \{b_i \neq b_j\})$
αντίστροφος ρόλος	R^{-}	$(R^{-})^{\mathcal{I}}(b, a) = R^{\mathcal{I}}(a, b)$
ισοδυναμία	$C \equiv D$	$\forall a \in \Delta^{\mathcal{I}}. C^{\mathcal{I}}(a) = D^{\mathcal{I}}(a)$
υπό-έννοια	$C \sqsubseteq D$	$\forall a \in \Delta^{\mathcal{I}}. C^{\mathcal{I}}(a) \leq D^{\mathcal{I}}(a)$
μεταβατικός ρόλος	$\text{Trans}(R)$	$\forall a, b \in \Delta^{\mathcal{I}}. R^{\mathcal{I}}(a, b) \geq \sup_{c \in \Delta^{\mathcal{I}}} \{t(R^{\mathcal{I}}(a, c), R^{\mathcal{I}}(c, b))\}$
υπό-ρόλος	$R \sqsubseteq S$	$\forall a, b \in \Delta^{\mathcal{I}}. R^{\mathcal{I}}(a, b) \leq S^{\mathcal{I}}(a, b)$
ισχυρισμοί εννοιών	$(a : C) \bowtie n$	$C^{\mathcal{I}}(a^{\mathcal{I}}) \bowtie n$
ισχυρισμοί ρόλων	$\langle (a, b) : R \rangle \bowtie n$	$R^{\mathcal{I}}(a^{\mathcal{I}}, b^{\mathcal{I}}) \bowtie n$

5.3.3 Κατανόηση Περιεχομένου και Συλλογιστική

Όπως είναι γνωστό, η διαδικασία της ανάλυσης εικόνας εμπεριέχει ένα μεγάλο ποσοστό αβεβαιότητας και ανακριβούς γνώσης, κάτι το οποίο οφείλεται καταρχήν στην υπόθεση που γίνεται ότι μία δισδιάστατη αναπαράσταση μπορεί να περιγράψει τον τρισδιάστατο κόσμο. Συνεπώς, είναι αναπόφευκτη η απώλεια πληροφορίας, για οποιαδήποτε περαιτέρω ανάλυση και αν θέλουμε να κάνουμε. Επομένως, είναι λογικό να υιοθετήσουμε ως μέσο πρακτικές της ασαφούς συνολοθεωρίας (όπως τις ορίσαμε στην ενότητα 2.2) και να χρησιμοποιήσουμε βαθμούς εμπιστοσύνης για το κατά πόσο μία περιοχή οπτικοποιεί μία έννοια, ή σε τι βαθμό ένα ζεύγος περιοχών ικανοποιεί ένα ρόλο. Έτσι μπορούμε να δηλώσουμε ότι μία περιοχή έχει τραχεία υφή σε βαθμό 0.7. Την τελευταία δεκαετία έχει ερευνηθεί η δυνατότητα ενσωμάτωσης της ασάφειας στις περιγραφικές λογικές [128, 125, 124] και έχουν δημιουργηθεί οι ασαφείς περιγραφικές λογικές (f-DLs).

Οι έννοιες, οι ρόλοι και τα αξιώματα στις ασαφείς περιγραφικές λογικές (fuzzy description logics - f-DLs) ορίζονται κατά τον ίδιο τρόπο που ορίστηκαν και στις κλασικές ΠΛ. Οι διαφορές των δύο γλωσσών εμφανίζονται στις σχέσεις μεταξύ ατόμων (ζεύγη ατόμων) και εννοιών (ρόλοι). Σε αντιδιαστολή με τις κλασικές ΠΛ, στα f-DLs ορίζουμε τους ασαφείς ισχυρισμούς (fuzzy assertions) που επιτρέπουν να περιγράψουμε σχέσεις στιγμιοτύπων (instance relations) χρησιμοποιώντας βαθμούς συμμετοχής. Πιο τυπικά, οι ασαφείς ισχυρισμοί ορίζονται ως εξής: $(a : C) \bowtie n$ ή $\langle (a, b) : R \rangle \bowtie n$, όπου \bowtie είναι $\geq, >, \leq$ και $<$. Μπορούμε να εξομοιώσουμε ισχυρισμούς της μορφής $(a : C) = n$, θεωρώντας δύο ισχυρισμούς της μορφής $(a : C) \geq n$ και $(a : C) \leq n$. Το σύνολο των ασαφών ισχυρισμών προσδιορίζει το ABox \mathcal{A} . Έτσι, η τριπλέτα των TBox, RBox και ABox $\langle \mathcal{T}, \mathcal{R}, \mathcal{A} \rangle$ είναι αυτό που αποκαλούμε οντολογία οπτικής πληροφορίας (visual ontology).

Οι ασαφείς ισχυρισμοί μπορούν να χρησιμοποιηθούν για να περιγράψουν τις σχέσεις στιγμιοτύπων μεταξύ μίας περιοχής και μίας έννοιας, αλλά και μεταξύ ενός ζεύγους περιοχών. Για παράδειγμα, μπορούμε να γράψουμε $(s_1 : \text{ΚΑΦΕ}) \geq 0.7$, $(\langle s_1, s_2 \rangle : \text{ΓΕΙΤΝΙΑΖΕΙ}) \geq 1$ και $(s_2 : \text{ΦΥΛΛΑ}) \geq 0.8$. Αυτή η γνώση που εξάγεται από ανάλυση της εικόνας, όπως θα δείχτει στην παράγραφο 5.4, αποτελεί τη ρητή γνώση (ενώ τα αξιώματα εννοιών και ρόλων αποτελούν την υπονοούμενη γνώση) και αρχικοποιεί την οπτική λογική.

Όπως αναφέραμε και παραπάνω, η οπτική λογική οφείλει να έχει τυπική σημασιολογία, επομένως οι ΠΛ είναι πράγματι κατάλληλα για την αναπαράσταση της οπτικής πληροφορίας. Η χρήση συναρτήσεων συμμετοχής με πεδίο τιμών το $[0, 1]$ εξυπηρετεί στον ορισμό της ασαφούς ερμηνείας (fuzzy interpretations) [128]. Μία ασαφής ερμηνεία είναι ένα ζεύγος $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, όπου το πεδίο $\Delta^{\mathcal{I}}$ είναι ένα μη-κενό σύνολο αντικειμένων και $\cdot^{\mathcal{I}}$ είναι μία συνάρτηση ασαφούς ερμηνείας, η οποία απεικονίζει:

1. ένα άτομο $a \in \mathcal{I}$ σε ένα στοιχείο $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$,
2. μία έννοια $C \in \mathcal{C}$ σε μία συνάρτηση συμμετοχής $C^{\mathcal{I}} : \Delta^{\mathcal{I}} \rightarrow [0, 1]$,
3. ένα ρόλο $R \in \mathcal{R}$ μία συνάρτηση συμμετοχής $R^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \rightarrow [0, 1]$.

Η συνάρτηση ασαφούς ερμηνείας μπορεί να επεκταθεί με χρήση τελεστών ασαφούς λογικής και να προσδώσει σημασιολογία σε σύνθετες έννοιες, ρόλους και αξιώματα. Ο πίνακας 5.1 απεικονίζει αυτή τη λογική, όπου c είναι το ασαφές συμπλήρωμα, t η ασαφής τομή και u η ασαφής ένωση (βλέπε παράγραφο 2.2.2).

Παράδειγμα 5.1 Έστω οι ασαφείς ισχυρισμοί $(s_1 : \text{ΚΑΦΕ}) \geq 0.8$ και $(s_1 : \text{ΤΡΑΧΥ}) \geq 0.9$ και το αξίωμα καθορισμού της έννοιας $\text{ΒΟΥΝΟ} \text{ ΒΟΥΝΟ} \equiv \text{ΚΑΦΕ} \sqcap \text{ΤΡΑΧΥ}$. Τότε, με βάση τον πίνακα 5.1 έχουμε $\text{ΚΑΦΕ}^{\mathcal{I}}(s_1^{\mathcal{I}}) \geq 0.8$ και $\text{ΤΡΑΧΥ}^{\mathcal{I}}(s_1^{\mathcal{I}}) \geq 0.9$ και μπορούμε να συμπεράνουμε ότι $\text{ΒΟΥΝΟ}^{\mathcal{I}}(s_1^{\mathcal{I}}) \geq t(\text{ΚΑΦΕ}^{\mathcal{I}}(\text{region}_1^{\mathcal{I}}), \text{ΤΡΑΧΥ}^{\mathcal{I}}(s_1^{\mathcal{I}})) \geq t(0.8, 0.9)$. Αν χρησιμοποιήσουμε την πρότυπη τομή για τ-νόρμα ($t(a, b) = \min(a, b)$), τότε $\text{ΒΟΥΝΟ}^{\mathcal{I}}(s_1^{\mathcal{I}}) \geq 0.8$, ενώ αν χρησιμοποιήσουμε την Lukasiewicz τ-νόρμα ($t(a, b) = \max(0, a + b - 1)$), τότε $\text{ΒΟΥΝΟ}^{\mathcal{I}}(s_1^{\mathcal{I}}) \geq 0.7$.

Μία ενδιαφέρουσα πλευρά των ΠΛ είναι ότι παρέχουν πλούσιες υπηρεσίες συμπερασματολογίας, για τις οποίες υπάρχουν αποφασίσιμοι (decidable) αλγόριθμοι. Αυτό σημαίνει ότι είναι εφικτό να εξαχθεί νέα γνώση με τεχνικές συλλογιστικής στην οντολογία. Οι υπηρεσίες συμπερασματολογίας των ΠΛ έχουν επεκταθεί και στις ασαφείς ΠΛ [128], ικανοποιώντας την απαίτηση για οπτική συλλογιστική. Επιπρόσθετα από τους τυπικούς μηχανισμούς συμπερασματολογίας των ΠΛ, οι ασαφείς ΠΛ παρέχουν την υπηρεσία υπολογισμού του μέγιστου ελάχιστου φράγματος (greatest lower bound - glb) ενός ατόμου σε μία έννοια, το οποίο ορίζεται ως εξής:

$$\text{glb}(a, C) = \sup\{n : \Sigma \models (a : C) \geq n\}, \quad (5.1)$$

όπου a είναι ένα άτομο, C μία έννοια, $n \in [0, 1]$, Σ η βάση γνώσης και \models παριστάνει τη λογική συνεπαγωγή (implication). Διαισθητικά, το glb είναι ο μεγαλύτερος βαθμός με τον οποίο ένα άτομο a ανήκει στην έννοια C χωρίς να εμφανίζεται κάποια αντίφαση στη γνώση. Όπως θα φανεί στη επόμενη ενότητα, η υπηρεσία συμπερασματολογίας glb θα φανεί ιδιαίτερα χρήσιμη για τη διαδικασία της ανάλυσης και κατανόησης της εικόνας.

Τέλος, προκειμένου να γίνει εφικτή η χρήση υπηρεσιών συλλογιστικής, είναι απαραίτητος ένας αποφασισμός αλγόριθμος για τη γλώσσα $f\text{-SHLN}$, όπως για παράδειγμα ο αλγόριθμος *tableaux* [126] για τη γλώσσα $f_{KD}\text{-SHLN}$. Η γλώσσα $f_{KD}\text{-SHLN}$ προκύπτει από την ασαφή περιγραφική γλώσσα $f\text{-SHLN}$ με το Lukasiewicz ασαφές συμπλήρωμα ($c(a) = 1 - a$), με τις Gödel τ-νόρμα και τ-κονόρμα ($t(a, b) = \min(a, b)$ και $u(a, b) = \max(a, b)$) και με την Kleene-Dienes ασαφή συνεπαγωγή ($\mathcal{J}(a, b) = \max(1 - a, b)$).

5.4 Από τους Οπτικούς Περιγραφείς στις Έννοιες-Διεγέρτες

Σε αυτή την ενότητα θα παρουσιάσουμε τη διαδικασία εξαγωγής των διεγερτών $T = \{\tau_i\}$ για περιοχές από τους οπτικούς περιγραφείς (κύρια αυτούς του προτύπου MPEG-7). Το σύνολο των διεγερτών $T = \{\tau_i\}$ μπορεί να αναλυθεί σε επιμέρους υποσύνολα T_i , που σχετίζονται με συγκεκριμένες ομάδες οπτικών περιγραφέων (π.χ. χρώμα, υφή, σχήμα) ή με εξειδικευμένους ανιχνευτές (π.χ. ανιχνευτής δέρματος) οι οποίοι προκύπτουν από ανάλυση πολλαπλών στοιχείων-χαρακτηριστικών (multiple cue): $T \equiv \bigcup_{i=1}^n T_i$.

Τα ασαφή σύνολα μπορούν να διαχειριστούν την αβεβαιότητα πολύ καλύτερα από τα κλασικά σύνολα, ενώ επιπλέον έχουν μεγαλύτερες εκφραστικές δυνατότητες για να αναπαραστήσουν προβλήματα του πραγματικού κόσμου [58]. Επομένως τα υιοθετούμε για την περιγραφή των ασαφών ισχυρισμών οι οποίοι αντιστοιχίζουν περιοχές της εικόνας και διεγέρτες με ένα βαθμό συμμετοχής. Ορίζουμε το ασαφές σύνολο των διεγερτών \mathcal{T} πάνω στο σύνολο T με συνάρτηση συμμετοχής $\mathcal{T}: T \rightarrow [0, 1]$. Χρησιμοποιώντας τη σημειογραφία αθροίσματος για ένα ασαφές σύνολο, γράφουμε

$$\mathcal{T} = \sum_{i=1}^{|T|} \tau_i / \mathcal{T}(\tau_i), \text{ όπου } |T| \text{ είναι η πληθυκότητα του συνόλου } T, \tau_i \in T \text{ είναι ένας}$$

διεγέρτης και $\mathcal{T}(\tau_i)$ είναι ο βαθμός συμμετοχής του διεγέρτη τ_i στο ασαφές σύνολο.

Στις επόμενες παραγράφους θα παρουσιάσουμε τους διεγέρτες που χρησιμοποιούμε και σχετίζονται με τρεις οπτικές πηγές: το χρώμα, την υφή και το σχήμα. Θα δείξουμε αναλυτικά τη διαδικασία αντιστοίχισης των οπτικών χαρακτηριστικών με τις έννοιες διεγέρτες και πως κατασκευάζουμε το ασαφές σύνολο \mathcal{T} .

5.4.1 Διεγέρτες Χρώματος

Το χρώμα αποτελεί ίσως το πιο εκφραστικό οπτικό χαρακτηριστικό, το οποίο έχει το επιπλέον πλεονέκτημα να είναι αμετάβλητο σε οποιοδήποτε γεωμετρικό μετασχηματισμό της εικόνας (μετατόπιση, περιστροφή, προβολή, κ.ά.). Στη βιβλιογραφία έχουν προταθεί πάρα πολλά χαρακτηριστικά που μπορούν να περιγράψουν το χρώμα μίας περιοχής της εικόνας. Ένας ευρέως διαδεδομένος περιγραφέας για το χρώμα είναι ο περιγραφέας κύριων χρωμάτων (dominant color - DC) [69], ο οποίος αποτελεί και μέρος του προτύπου MPEG-7. Για μία περιοχή s_1 , ο περιγραφέας DC ορίζεται ως ένα διάνυσμα χαρακτηριστικών:

$$DC \equiv \langle (c, p, v)_i, s \rangle, \text{ όπου } i = 1, \dots, N \quad (5.2)$$

Στην παραπάνω παράσταση, c_i είναι το i -οστό κύριο χρώμα, v_i είναι η διακύμανση της κατανομής του χρώματος στην περιοχή, p_i είναι το ποσοστό κατάληψης της περιοχής από το i -οστό κύριο χρώμα, s είναι η χωρική συνοχή στην έκταση της περιοχής και N είναι ο αριθμός κύριων χρωμάτων που μπορεί να είναι το μέγιστο έως οκτώ. Ωστόσο, η αναπαράσταση αυτή δεν είναι κατάλληλη για τη δική μας προσέγγιση, καθώς δεν επιθυμούμε να χρησιμοποιήσουμε το συγκεκριμένο περιγραφέα απευθείας για κάποια εφαρμογή, όπως π.χ. για δεικτοδότηση και ανάκτηση εικόνων με βάση το περιεχόμενο (όπου θα ορίζαμε κάποια μετρική απόστασης πάνω στο διάνυσμα DC). Σκοπός μας είναι να αντιστοιχίσουμε τις αριθμητικές τιμές του διανύσματος DC με γλωσσικούς όρους, οι οποίοι θα αποτελούν τους διεγέρτες (ορισμός 5.1, παράγραφος 5.3) για το χρώμα, οι οποίοι είναι:

$$T_1 = \{\text{ΚΟΚΚΙΝΟ, ΑΝΟΙΧΤΟΚΟΚΚΙΝΟ, ΚΙΤΡΙΝΟ, ΑΝΟΙΧΤΟΚΙΤΡΙΝΟ, ΠΡΑΣΙΝΟ, ΑΝΟΙΧΤΟΠΡΑΣΙΝΟ, ΜΠΛΕ, ΑΝΟΙΧΤΟΜΠΛΕ, ΜΑΥΡΟ, ΓΚΡΙ, ΑΝΟΙΧΤΟΓΚΡΙ, ΑΣΠΡΟ}\}$$

Για να επιτύχουμε να χαρακτηρίσουμε με γλωσσικούς όρους αριθμητικές τιμές, χρειαζόμαστε μία συνάρτηση μετασχηματισμού που να συνδέει το διάνυσμα DC με ένα (ασαφές) σύνολο ευδιάκριτων χρωμάτων, όπως αυτά που χρησιμοποιούμε καθημερινά. Τα χρωματικά σύνολα (color sets) [119] είναι μία συμπαγής και πιο συμπυκνωμένη μορφή αναπαράστασης των ιστογραμμάτων χρώματος και έχουν προταθεί διάφοροι μέθοδοι για την εξαγωγή τους, όπως η κατωφλίωση επί του ιστογράμματος ή επί της πίσω-προβολής (back-projection) της εικόνας. Στη δική μας περίπτωση, το χρωματικό σύνολο αποτελείται από το σύνολο των διεγερτών T_1 και κατασκευάζεται με βάση τον περιγραφέα κύριων χρωμάτων. Επιπλέον, για κάθε διεγέρτη υπολογίζουμε και το βαθμό που εκφράζει πόσο καλά χαρακτηρίζει ο γλωσσικός όρος τις αριθμητικές τιμές. Αυτός ο βαθμός είναι ο βαθμός συμμετοχής του διεγέρτη (τ_i) στο

$$\text{ασαφές σύνολο } T_1 = \sum_{i=1}^{|T_1|} \tau_i / T_1(\tau_i), \text{ όπου } \tau_i \in T_1.$$

Η διαδικασία κατασκευής του ασαφούς χρωματικού συνόλου, δεδομένου του περιγραφέα κύριων χρωμάτων, είναι η εξής: Πρώτα μετατρέπουμε το διάνυσμα χρώματος $\mathbf{v} = [r \ g \ b]$ από τον RGB χρωματικό χώρο στο διάνυσμα $\mathbf{u} = [h \ s \ v]$ του HSV , ο οποίος είναι οπτικά πιο ομοιόμορφος (περισσότερες λεπτομέρειες αναφέρονται στην παράγραφο 3.4.2.

Το επόμενο βήμα είναι να χβαντίσουμε το \mathbf{u} και θα το κάνουμε σε 12 επίπεδα: $q[m] = Q(\mathbf{u})$, όπου $q[m]$ είναι το διάνυσμα χβαντισμού και $m \in [1, \dots, 12]$. Η συνιστώσα H (Hue) που περιγράφει τη χρωματική απόχρωση χβαντίζεται σε 4 επίπεδα, η συνιστώσα S (saturation) που περιγράφει τον κορεσμό του χρώματος χβαντίζεται σε 1 επίπεδο, η συνιστώσα V που περιγράφει τη φωτεινότητα σε δύο επίπεδα (ανοιχτό και σκούρο) και τέλος άλλα 4 επίπεδα της φωτεινότητας με μηδενική τιμή κορεσμού αντιστοιχούν σε 4 τιμές γκρι, από μαύρο έως άσπρο. Το διάνυσμα χβαντισμού $q[m]$ καθορίζει έναν 12-διάστατο δυαδικό χώρο B^{12} , όπου κάθε διεγέρτης χρώματος αντιστοιχεί σε μία επιλογή χρώματος από τον χβαντισμένο αυτό χρωματικό χώρο.

Η ασάφεια εισάγεται στο βήμα χβαντισμού της χρωματικής απόχρωσης H . Ο χρωματικός χώρος HSV είναι ένας κυλινδρικός χώρος, του οποίου η συνιστώσα H λαμβάνει τιμές στο πεδίο $[0, 360^\circ]$. Ο χβαντισμός σε 4 στάθμες του H γίνεται με το διαχωρισμό του δίσκου του σε 4 τομείς που αντιστοιχούν σε 4 διακριτές χρωματικές αποχρώσεις: κόκκινο, κίτρινο, πράσινο και μπλε, με τα κέντρα τους στις θέσεις 0° , 90° , 180° και 150° , αντιστοίχως. Για κάθε ένα, ορίζουμε μία συνάρτηση συμμετοχής



Σχήμα 5.3: Συναρτήσεις συμμετοχής για τα χρώματα κόκκινο, κίτρινο, πράσινο και μπλε δειγματοληπτούμενα στο δίσκο της χρωματικής απόχρωσης H του HSV χώρου

$\mathcal{T}_1(\tau_i): [0, 2\pi) \rightarrow [0, 1]$ που αντιστοιχίζει τις τιμές της απόχρωσης θ του διεγέρτη τ_i στο μοναδιαίο διάστημα. Η εξίσωση (5.3) παριστά τη συνάρτηση συμμετοχής για το διεγέρτη ΚΟΚΚΙΝΟ· παρόμοια ορίζονται οι συναρτήσεις συμμετοχής και για τους υπόλοιπους χρωματικούς διεγέρτες. Καθώς η απόχρωση ορίζεται σε πολικές συντεταγμένες, δικαιολογείται να χρησιμοποιήσουμε την προβολή στους δύο άξονες ως συνάρτηση συμμετοχής. Το σχήμα 5.3 δείχνει τις συναρτήσεις συμμετοχής για τις τέσσερις χρωματικές αποχρώσεις κόκκινο, κίτρινο, πράσινο και μπλε.

$$\mathcal{T}_1(\text{ΚΟΚΚΙΝΟ}) = \begin{cases} \cos(\theta), & \theta \in [0, \frac{\pi}{2}) \\ 0, & \theta \in [\frac{\pi}{2}, \frac{3\pi}{2}) \\ \cos(\theta), & \theta \in [\frac{3\pi}{2}, 2\pi). \end{cases} \quad (5.3)$$

Συνδυάζοντας τα 4 επίπεδα της απόχρωσης (κόκκινο, κίτρινο, πράσινο, μπλε) με τα δύο της φωτεινότητας (σκούρο και ανοιχτό) καταλήγουμε σε οκτώ διεγέρτες. Οι υπόλοιποι τέσσερις εξάγονται με ομοιόμορφη κβάντιση της φωτεινότητας με ταυτόχρονη μηδενική χρωματική απόχρωση, παράγοντας κατ' αυτόν τον τρόπο τέσσερις επιπλέον διεγέρτες: Μαύρο, Γκρι, ΑνοιχτόΓκρι και Άσπρο.

5.4.2 Διεγέρτες Υφής

Ο περιγραφέας αναζήτησης υφής (texture browsing - TB) χαρακτηρίζει την υφή σε σχέση με την κανονικότητα (regularity), την κατεύθυνση (directionality) και τραχύτητα (coarseness) [69]. Ο περιγραφέας TB ορίζεται στο πρότυπο MPEG-7 ως εξής:

$$TB \equiv [v_1 \ v_2 \ v_3 \ v_4 \ v_5]. \quad (5.4)$$

Στο παραπάνω διάγραμμα το v_1 αντιπροσωπεύει το βαθμό κανονικότητας της υφής και είναι κβαντισμένος σε τέσσερις τιμές: πολύ κανονικό, κανονικό, λίγο ακανόνιστο και ακανόνιστο. Οι τιμές v_2 και v_3 ρυθμίζουν τις δύο κύριες κατευθύνσεις (προσανατολισμούς) της υφής. Σημειώνεται ότι μία εικόνα μπορεί να έχει ταυτόχρονα μέχρι το πολύ δύο κατευθύνσεις, π.χ. ένας τοίχος αποτελούμενος από τούβλα έχει την ίδια στιγμή και οριζόντια και κατακόρυφη κατευθυντική υφή. Η περιγραφή της κατεύθυνσης κβαντίζεται σε έξι επίπεδα, στο πεδίο τιμών $[0, 150^\circ]$, ενώ υπάρχει επιπλέον και η τιμή της μη κατεύθυνσης. Τέλος οι τιμές v_4, v_5 αντιστοιχούν στην τραχύτητα, η οποία κβαντίζεται σε τέσσερα επίπεδα: λείο, λίγο τραχύ, τραχύ και πολύ τραχύ. Περισσότερες λεπτομέρειες για τον περιγραφέα αναζήτησης υφής και την εξαγωγή του από μία εικόνα ή περιοχή της εικόνας δίνονται στο άρθρο [143].

Πίνακας 5.2: Ορισμός βαθμών συμμετοχής διεγερτών υψής.

Κανονικότητα (v_1)	T_2 (καν)	Τραχύτητα (v_4 & v_5)	T_2 (τρχ)	Κατεύθυνση (v_2 & v_3)	T_2 (κτθ)
ακανόνιστο	0	λείο	0	οριζόντια	$\max(\cos(v_2), \cos(v_3))$
λίγο ακανόνιστο	0.3	λίγο τραχύ	0.3	κατακόρυφη	$\max(\sin(v_2), \sin(v_3))$
κανονικό	0.8	τραχύ	0.8		
πολύ κανονικό	1	πολύ τραχύ	1		

Ο περιγραφέας TB έχει από τη φύση του ήδη μία σημασιολογική έννοια, καθώς δεν περιγράφεται από αριθμητικές τιμές, αλλά από γλωσσικούς όρους. Αυτός είναι και ένας βασικός λόγος για τον οποίο επιλέξαμε αυτόν τον περιγραφέα, ο οποίος δημιουργήθηκε στο πλαίσιο του προτύπου MPEG-7 για την αναζήτηση παρόμοιων σε υφή εικόνων. Οι γλωσσικοί αυτοί όροι είναι πάντως έτσι ορισμένοι, ώστε αυστηρά είτε υπάρχουν είτε όχι, κάτι το οποίο είναι χρήσιμο για εφαρμογές δεικτοδότησης και αναζήτησης. Στη δική μας όμως αρχιτεκτονική, επιθυμούμε να εισάγουμε ασάφεια στους όρους αυτούς και στις ιδιότητες αυτές της υψής (κανονικότητα, κατεύθυνση και τραχύτητα) και να προσδώσουμε ένα βαθμό συμμετοχής. Για αυτόν το σκοπό, ορίζουμε μία συνάρτηση μετασχηματισμού που απεικονίζει τις τιμές του περιγραφέα TB σε διεγέρτες υψής σε ένα βαθμό. Το σύνολο αυτών των διεγερτών είναι:

$$T_2 = \{ \text{ΤΡΑΧΥΤΗΤΑ, ΚΑΝΟΝΙΚΟΤΗΤΑ, ΟΡΙΖΟΝΤΙΑΚΑΤΕΥΘΥΝΣΗ, ΚΑΤΑΚΟΡΥΦΗΚΑΤΕΥΘΥΝΣΗ} \}$$

Με πρώτη ματιά κάποιος μπορεί να πιστευε ότι το σύνολο των διεγερτών υψής T_2 αποτελεί υποσύνολο του περιγραφέα TB , αυτό ισχύει μόνο για το κλασικό σύνολο T_2 και όχι για το ασαφές σύνολο \mathcal{T}_2 . Το ασαφές σύνολο \mathcal{T}_2 παρέχει το απαραίτητο εύρος τιμών για τα στοιχεία του, ώστε με το βαθμό συμμετοχής να εκφράζεται άμεσα το πόσο αντιπροσωπευτικός είναι ένας διεγέρτης για μία περιοχή της εικόνας: $T_2 = \sum_{i=1}^{|\mathcal{T}_2|} \tau_i / \mathcal{T}_2(\tau_i)$, όπου $\tau_i \in \mathcal{T}_2$

Η διαδικασία ασαφοποίησης των διεγερτών υψής, είναι μία απεικόνιση των βαθμών του περιγραφέα TB σε βαθμούς συμμετοχής, καθώς ο TB ορίζεται στο διακριτό πεδίο (πίνακας 5.2). Στο διεγέρτη υψής ΚΑΝΟΝΙΚΟΤΗΤΑ, αναθέτουμε τέσσερις βαθμούς συμμετοχής, ανάλογα με την τιμή v_1 του περιγραφέα TB (εξίσωση (5.4)). Παρόμοια, για το διεγέρτη υψής ΤΡΑΧΥΤΗΤΑ αναθέτουμε τέσσερις πάλι τιμές, ανάλογα με την ένταση της τραχύτητας των v_2 και v_3 . Όσον αφορά την κατεύθυνση της υψής, ορίζουμε τους εξής δύο διεγέρτες: ΟΡΙΖΟΝΤΙΑΚΑΤΕΥΘΥΝΣΗ και ΚΑΤΑΚΟΡΥΦΗΚΑΤΕΥΘΥΝΣΗ. Οι βαθμοί συμμετοχής προκύπτουν από την προβολή της κάθε κύριας κατεύθυνσης (μία από τις έξι που υποστηρίζει ο TB) στον οριζόντιο και τον κατακόρυφο άξονα αντιστοίχως. Οπότε, οι συναρτήσεις συμμετοχής είναι: $\mathcal{T}_2(\text{ορζ}) = \max(\cos(v_2), \cos(v_3))$ $\mathcal{T}_2(\text{κατ}) = \max(\sin(v_2), \sin(v_3))$.

5.4.3 Διεγέρτες Σχήματος

Για την περιγραφή του σχήματος μίας περιοχής δεν θα χρησιμοποιήσουμε κάποιον από τους περιγραφείς του προτύπου MPEG-7, καθώς οι διεγέρτες που θέλουμε να ορίσουμε προκύπτουν απευθείας από την εικόνα. Το σύνολο, λοιπόν, των διεγερτών σχήματος είναι το εξής:

$$T_3 = \{ \text{ΚΥΚΛΙΚΟΤΗΤΑ, ΟΡΘΟΓΩΝΙΟΤΗΤΑ, ΕΠΙΜΗΚΥΝΣΗ} \}$$

Η κυκλικότητα είναι ένα μέτρο του κατά πόσο κυκλική και συμπαγής είναι μία περιοχή και υπολογίζεται ως εξής:

$$\text{κυκλικότητα} = \frac{4\pi \cdot \text{εμβαδόν}}{\text{περίμετρος}^2} \quad (5.5)$$

Πρόκειται για ένα μέγεθος χωρίς μονάδες, του οποίου η τιμή ελαχιστοποιείται, όπως αναμένεται, για μία κυκλική περιοχή, ενώ όσο μεγαλύτερη είναι η τιμή του τόσο λιγότερο κυκλικό είναι το σχήμα της περιοχής. Σύμφωνα με αυτή την παρατήρηση, επιλέγουμε μία συνεχή, γνησίως φθίνουσα συνάρτηση για την απεικόνιση της κυκλικότητας σε βαθμό συμμετοχής στο ασαφές σύνολο του διεγέρτη ΚΥΚΛΙΚΟΤΗΤΑ:

$$T_3(\text{κυκλ}) = e^{-1/\text{κυκλ}}: [0, 1] \rightarrow [0, 1] \quad (5.6)$$

Η ορθογωνιότητα είναι ένα μέγεθος που δείχνει πόσο καλά ταιριάζει η υπό εξέταση περιοχή στο περικλείον πλαίσιο (bounding box) και υπολογίζεται με βάση την παρακάτω σχέση:

$$\text{ορθογωνιότητα} = \frac{\text{εμβαδόν περιοχής}}{\text{εμβαδόν περικλείοντος ορθογωνίου}} \quad (5.7)$$

Προφανώς, για ορθογώνιες περιοχές, οι οποίες ταυτίζονται με το περικλείον τους ορθογώνιο, η τιμή της ορθογωνιότητας είναι ίση με τη μονάδα, ενώ όσο μειώνεται η τιμή αυτή, τόσο λιγότερο ορθογώνια είναι η περιοχή. Με βάση αυτή την παρατήρηση, ορίζουμε ως συνάρτηση συμμετοχής του διεγέρτη ΟΡΘΟΓΩΝΙΟΤΗΤΑ την ταυτοτική συνάρτηση:

$$T_3(\text{ορθογ}) = \text{ορθογ}: [0, 1] \rightarrow [0, 1] \quad (5.8)$$

Τέλος, η επιμήκυνση ή εκκεντρότητα, είναι ένα ακόμα χαρακτηριστικό μέγεθος για το σχήμα μίας περιοχής και ισούται με το λόγο του μήκους της μέγιστης χορδής που διατρέχει την περιοχή, προς το μήκος της ελάχιστης χορδής:

$$\text{επιμήκυνση} = \frac{\text{μήκος μέγιστης χορδής}}{\text{μήκος ελάχιστης χορδής}} \quad (5.9)$$

Διαισθητικά, η συνάρτηση συμμετοχής για αυτόν το διεγέρτη, πρέπει να είναι γνησίως αύξουσα, και μάλιστα με μη-γραμμικό τρόπο, μιας και η ιδιότητα της επιμήκυνσης υπολογίζεται ως κλάσμα δύο μεταβλητών. Έτσι προτείνουμε την παρακάτω συνάρτηση συμμετοχής για τον υπολογισμό του βαθμού συμμετοχής του διεγέρτη ΕΠΙΜΗΚΥΝΣΗ:

$$T_3(\text{επιμ}) = 1 - 1/\text{επιμ}: [1, \infty) \rightarrow [0, 1] \quad (5.10)$$

Έχοντας ορίσει το σύνολο T_3 των διεγερτών σχήματος και τη συνάρτηση συμμετοχής για τον καθένα, μπορούμε να ορίσουμε το ασαφές σύνολο των διεγερτών

$$\text{σχήματος } T_3 = \sum_{i=1}^{|T_3|} \tau_i / T_3(\tau_i), \text{ όπου } \tau_i \in T_3.$$

Πίνακας 5.3: Δείγμα της βάσης γνώσης Σ .

$\mathcal{T} = \{$	Γ	\equiv	ΑΝΟΙΧΤΟΠΡΑΣΙΝΟ,
	Φ	\equiv	ΠΡΑΣΙΝΟ \sqcup ΚΑΝΟΝΙΚΗΥΦΗ,
	ΒΟΥΝΟ	\equiv	(ΠΡΑΣΙΝΟ \sqcup ΚΑΦΕ) \sqcap ΤΡΑΧΕΙΑΥΦΗ,
	ΧΩΜΑΤΙΝΟΣΔΡΟΜΟΣ	\equiv	(ΚΑΦΕ \sqcup ΑΝΟΙΧΤΟΚΑΦΕ) \sqcap ΤΡΑΧΕΙΑΥΦΗ,
	ΑΣΦΑΛΤΙΝΟΣΔΡΟΜΟΣ	\equiv	(ΓΚΡΙ \sqcup ΑΝΟΙΧΤΟΓΚΡΙ) \sqcap ΛΕΙΟ,
	ΛΑΣΠΩΔΗΣΔΡΟΜΟΣ	\equiv	ΚΑΦΕ \sqcap ΤΡΑΧΕΙΑΥΦΗ,
	ΧΙΟΝΙΣΜΕΝΟΣΔΡΟΜΟΣ	\equiv	ΑΣΠΡΟ \sqcap ΛΕΙΟ,
	Δ	\equiv	$\text{ΧΩΜΑΤΙΝΟΣΔΡΟΜΟΣ} \sqcup \text{ΑΣΦΑΛΤΙΝΟΣΔΡΟΜΟΣ} \sqcup$ $\text{ΛΑΣΠΩΔΗΣΔΡΟΜΟΣ} \sqcup \text{ΧΙΟΝΙΣΜΕΝΟΣΔΡΟΜΟΣ},$
	Δ	\equiv	$\text{ΚΑΦΕ} \sqcap \text{ΞΓΕΙΤΝΙΑΖΕΙ.}\Phi\text{ΥΛΛΩΜΑ}\}$

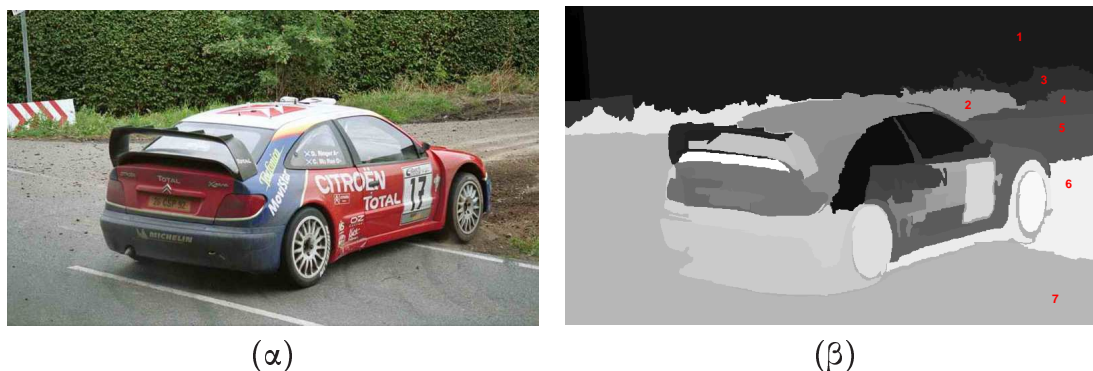
$\mathcal{R} = \{\text{ΓΕΙΤΝΙΑΖΕΙ} \sqsubseteq \text{ΓΕΙΤΝΙΑΖΕΙ}^-\}$

5.5 Αρχικά Αποτελέσματα

Όλη η διαδικασία της ανάλυσης ξεκινάει με την κατασκευή του αρχικού σώματος ισχυρισμών ABox , όπου οι αρχικοί ισχυρισμοί συνιστώνται αποκλειστικά από έννοιες-διεγέρτες. Επίκληση της οπτικής συλλογιστικής χρησιμοποιεί το αρχικό ABox και τα αξιώματα του σώματος ορολογίας TBox και δημιουργεί νέους ισχυρισμούς, συνδέοντας περιοχές της εικόνας με νέες έννοιες. Επιπλέον, με χρήση της υπηρεσίας συλλογιστικής μέγιστου ελάχιστου φράγματος glb μπορούμε να ανακτήσουμε πληροφορίες για το βαθμό ομοιότητας δύο γειτονικών περιοχών, σε σχέση με κάποια έννοια, και να τις συγχωνεύσουμε σε μία ενιαία περιοχή η οποία αναπαριστά την ίδια έννοια.

Στο κεφάλαιο 5.3 ορίσαμε τη δομή και τη λειτουργία της προτεινόμενης οπτικής λογικής. Αυτή η λογική αποσκοπεί στο να μας βοηθήσει να αντιληφθούμε τις έννοιες μίας θεματικής περιοχής καθώς και τις σχέσεις που τις διέπουν, με τρόπο που να εξυπηρετεί αλγορίθμους ανάλυσης εικόνας και να τους παρέχει το υπόβαθρο γνώσης για την επίτευξη ποιοτικών και σημασιολογικών αποτελεσμάτων. Στον πίνακα 5.3 εμφανίζεται ένα απόσπασμα της οντολογίας του θεματικού πεδίου “ρόλι”. Πρόκειται για ένα πολύ μικρό μέρος της βάσης γνώσης Σ , κάτι που έγινε εσκεμμένα για να είναι πιο κατανοητή στον αναγνώστη η διαδικασία της συλλογιστικής και να γίνει πιο δυνατή η παρακολούθηση της βήμα προς βήμα. Πρώτα θα δείξουμε ότι με χρήση των εννοιών-διεγερτών που ορίσαμε στην ενότητα 5.4 μπορούμε να κατασκευάσουμε απλά αξιώματα. Αυτά τα αξιώματα χρησιμοποιούν τις έννοιες-διεγέρτες μαζί με μερικούς κατασκευαστές ΠΛ με σκοπό τον ορισμό απλών εννοιών.

Παραδείγματος χάριν, έχουμε ορίσει την έννοια Γ χρησιμοποιώντας ένα διεγέρτη χρώματος, συγκεκριμένα τον ΑΝΟΙΧΤΟΠΡΑΣΙΝΟ . Επίσης, έχουμε ορίσει την έννοια ΛΑΣΠΩΔΗΣΔΡΟΜΟΣ συνδυάζοντας δύο διεγέρτες, έναν χρώματος και έναν υφής. Από την άλλη, επιτρέποντας χρήση ρόλων, μπορούμε να συνδυάσουμε περιοχές, με διαφορετικά μάλιστα οπτικά χαρακτηριστικά, για να ορίσουμε πιο σύνθετες έννοιες, αξιοποιώντας δομικά αξιώματα (ορισμός 5.3). Πιο συγκεκριμένα, ο ρόλος ΓΕΙΤΝΙΑΖΕΙ χρησιμοποιείται για να ορίσουμε τη σύνθετη έννοια Δ , το οποίο χοντρικά αποτελείται από δύο κομμάτια, τον κορμό και το φύλλωμα. Τέλος, στον πίνακα 5.3 παρουσιάζουμε και ένα αξίωμα ρόλου το οποίο δηλώνει ότι ο ρόλος



Σχήμα 5.4: Αρχική εικόνα (αριστερά) και η κατάτμησή της τη χρονική στιγμή t (δεξιά)

ΓΕΙΤΝΙΑΖΕΙ είναι συμμετρικός. Με τη βοήθεια της οπτικής λογικής μπορούμε να κατασκευάσουμε περισσότερους ορισμούς εννοιών, απλούστερων και συνθετότερων, όπως είναι οι έννοιες ΟΥΡΑΝΟΣ, ΣΥΝΝΕΦΙΑΣΜΕΝΟΣ ΟΥΡΑΝΟΣ, ΡΟΔΑ και ΑΥΤΟΚΙΝΗΤΟ.

Στο σχήμα 5.4 απεικονίζεται μία εικόνα και δίπλα μία πιθανή κατάτμησή της σε μία φάση t του αλγορίθμου κατάτμησης, όπου έχει χωριστεί σε περίπου τριάντα περιοχές. Σε αυτό το σημείο ας επανέλθουμε λίγο στον αλγόριθμο κατάτμησης *SRSST* που παρουσιάσαμε αναλυτικά στην παράγραφο 4.2.3. Κατά την εκτέλεση του t -οστού βήματος συγχώνευσης περιοχών, έχουμε ήδη ένα σώμα ισχυρισμών που έχει προκύψει από τον εντοπισμό των εννοιών-διεγερτών στα προηγούμενα $(t - 1)$ βήματα, καθώς και στην εξαγωγή εννοιών με βάση την οπτική συλλογιστική. Ο πίνακας 5.4-A παρέχει το ασαφές σύνολο ισχυρισμών για ένα συγκεκριμένο κομμάτι της εικόνας (οι κωδικοί αριθμοί των περιοχών αντιστοιχούν με αυτούς της κατατμημένης εικόνας του σχήματος 5.4). Σημειώνεται ότι έννοιες-διεγέρτες που δεν έχουν ανιχνευθεί (ή ισοδύναμα, εντοπισμός διεγέρτη με μηδενικό βαθμό συμμετοχής στο ασαφές σύνολο των διεγερτών) δεν παράγει ισχυρισμό του τύπου $(s_i \text{ ΕΝΝΟΙΑ}_j \geq 0)$, καθώς δεν προσφέρουν κάποια σημασιολογική αξία στη διαδικασία της οπτικής συλλογιστικής, ενώ αντιθέτως αυξάνουν την πολυπλοκότητα και μειώνουν την απόδοση του συστήματος.

Επίκληση της οπτικής συλλογιστικής δημιουργεί νέους ασαφείς ισχυρισμούς (βλέπε πίνακα 5.4-B) εμπλουτίζοντας το *ABox*, ενώ επιπλέον επαναυπολογίζεται το μέγιστο ελάχιστο φράγμα *glb* για κάθε περιοχή και για κάθε έννοια σύμφωνα με την εξίσωση (4.6). Για παράδειγμα, έστω η περιοχή s_1 στην οποία έχει αποδοθεί η έννοια ΦΥΛΛΑ σύμφωνα με τον αξιωματικό ορισμό: $\text{ΦΥΛΛΑ} \equiv \text{ΠΡΑΣΙΝΟ} \sqcap \text{REGULARTEXTURED}$, όπου ο βαθμός για τη νέα έννοια προκύπτει από την ασαφή ένωση των βαθμών των σχετιζόμενων διεγερτών. Σαρώνοντας το γράφο *ARG*, βρίσκουμε ότι οι περιοχές s_2 και s_3 έχουν μέγιστη ομοιότητα και άρα η ακμή e_{23} του *ARG* έχει ελάχιστο βάρος w_{23} , όπως αυτό υπολογίζεται από την εξίσωση (4.5). Οι δύο αυτές περιοχές σε αυτό το βήμα του αλγορίθμου, οπτικοποιούν από κοινού την έννοια ΓΡΑΣΙΔΙ· τονίζουμε ότι αυτό ισχύει στο t -βήμα, καθώς σε επόμενη φάση μπορούν κάλλιστα να οπτικοποιούν κάποια άλλη έννοια, όπως π.χ. την ΚΗΠΟΣ. Με βάση τη σημασιολογική τους ομοιότητα και την από κοινού οπτικοποίηση της έννοιας ΓΡΑΣΙΔΙ, αποφασίζουμε να τις συγχωνεύσουμε και να επαναλάβουμε τη διαδικασία εξαγωγής διεγερτών για τη νέα περιοχή. Οι νέοι ισχυρισμοί προστίθενται στο *ABox* και η διαδικασία προχωράει στο βήμα $(t + 1)$.

Πίνακας 5.4: (α) Ασαφείς ισχυρισμοί από ένα δείγμα του *ABox* στο βήμα *t* και (β) οι εξαγόμενοι ασαφείς ισχυρισμοί.

Περιοχή	(α)		(β)	
	Έννοια	Βαθμός	Έννοια	Βαθμός
s_1	ΠΡΑΣΙΝΟ	0.79	ΦΥΛΛΩΜΑ	0.79
s_1	ΤΡΑΧΕΙΑΥΦΗ	0.80	ΓΡΑΣΙΔΙ	0.78
s_1	ΚΑΝΟΝΙΚΗΥΦΗ	0.80	ΓΡΑΣΙΔΙ	0.71
s_1	ΟΡΘΟΓΩΝΙΟΥΣΧΗΜΑΤΟΣ	0.87	ΒΟΥΝΟ	0.69
s_2	ΑΝΟΙΧΤΟΠΡΑΣΙΝΟ	0.78	ΧΩΜΑΤΙΝΟΣΔΡΟΜΟΣ	0.69
s_2	ΤΡΑΧΕΙΑΥΦΗ	0.80	ΛΑΣΠΩΔΗΣΔΡΟΜΟΣ	0.69
s_3	ΑΝΟΙΧΤΟΠΡΑΣΙΝΟ	0.71	ΔΡΟΜΟΣ	0.69
s_3	ΤΡΑΧΕΙΑΥΦΗ	0.30	ΧΩΜΑΤΙΝΟΣΔΡΟΜΟΣ	0.30
s_4	ΚΑΦΕ	0.69	ΔΡΟΜΟΣ	0.30
s_4	ΤΡΑΧΕΙΑΥΦΗ	0.80	ΣΚΟΝΗ	0.70
s_5	ΑΝΟΙΧΤΟΠΡΑΣΙΝΟ	0.85	ΒΟΥΝΟ	0.67
s_5	ΤΡΑΧΕΙΑΥΦΗ	0.30	ΧΩΜΑΤΙΝΟΣΔΡΟΜΟΣ	0.67
s_6	ΚΑΦΕ	0.67	ΛΑΣΠΩΔΗΣΔΡΟΜΟΣ	0.67
s_6	ΤΡΑΧΕΙΑΥΦΗ	0.80	ΔΡΟΜΟΣ	0.67
s_7	ΑΝΟΙΧΤΟΓΚΡΙ	0.72	ΑΣΦΑΛΤΙΝΟΣΔΡΟΜΟΣ	0.70
s_7	ΤΡΑΧΕΙΑΥΦΗ	0.30	ΔΡΟΜΟΣ	0.70

5.6 Συναφείς Εργασίες

Η ιδέα της χρήσης γλώσσας τυπικής αναπαράστασης γνώσης σε προβλήματα ανάλυσης πολυμέσων δεν είναι καινούρια. Στο παρελθόν έχουν χρησιμοποιηθεί ασαφείς περιγραφικές λογικές (f-DLs) προκειμένου να παρέχουν ένα τυπικό μοντέλο για ανάκτηση πολυμεσικής πληροφορίας [75, 76]. Μετέπειτα εργασία του Di Sciascio [108, 33] παρουσίασε μία επέκταση των ΠΛ ώστε να παρέχουν τα μέσα να αξιοποιηθεί το περίγραμμα και το χρώμα μίας περιοχής ως έννοιες περιγραφικών λογικών και το προτεινόμενο μοντέλο χρησιμοποιήθηκε για ανάκτηση εικόνων με βάση το σχήμα. Οι Möller και Neumann [80, 90] παρουσίασαν τη δική τους προσέγγιση για την κατανόηση εικόνων με βάση τυπικές περιγραφικές γλώσσες. Στην πρώτη εργασία [80] διερευνήθηκαν οι επιπλέον δυνατότητες που παρέχουν οι πρότυπες επεκτάσεις των ΠΛ [11], ενώ στη δεύτερη εργασία [90] χρησιμοποιήθηκαν και επεκτάσεις των ΠΛ για χωρικές σχέσεις. Όλες οι προαναφερθείσες προσεγγίσεις θεωρούν δεδομένη την ύπαρξη ορθής και εκφραστικής περιγραφής της εικόνας με βάση ετικέτες που έχουν αντιστοιχηθεί σε περιοχές της, δηλαδή προϋποθέτουν την εκτέλεση ενός αλγορίθμου κατάτμησης και εν συνεχεία μία διαδικασία αναγνώρισης που παρέχει την αρχική περιγραφή η οποία αποθηκεύεται σε μία βάση δεδομένων. Οι περιγραφικές λογικές χρησιμοποιούνται αποκλειστικά για να υλοποιήσουν εργασίες υψηλότερου σημασιολογικού επιπέδου ή για ανάκτηση περιεχομένου αποθηκευμένου σε μία βάση γνώσης. Καθ' όσον είμαστε σε θέση να γνωρίζουμε, η δική μας προσέγγιση διαφοροποιείται και είναι μοναδική στο γεγονός ότι χρησιμοποιεί τις ΠΛ και υπηρεσίες συλλογιστικής αυτών σε προβλήματα ανάλυσης εικόνας χαμηλού επιπέδου, όπως είναι η κατάτμηση, με σκοπό να βελτιώσει τόσο τα αποτελέσματα της κατάτμησης, όσο και της αναγνώρισης αντικειμένων. Η καινοτομία της προτεινόμενης μεθοδολογίας είναι η στενή διασύνδεση

μεθόδων που εφαρμόζονται απευθείας στην εικόνα και τις τιμές των εικονοστοιχείων της, με τεχνολογίες που βασίζονται σε διαχείριση γνώσης υψηλού επιπέδου και υπηρεσίες συλλογιστικής· κατ' αυτόν τον τρόπο οι διαδικασίες της κατάτμησης και της αναγνώρισης έρχονται πιο κοντά.

Επίσης, έχουν προταθεί και ορισμένες τεχνικές που βασίζονται σε τεχνολογίες άλλες από τις περιγραφικές λογικές. Η Hunter [53] πρότεινε τη χρήση οντολογιών και κανόνων. Πιο συγκεκριμένα, οι κανόνες κατασκευάζονται με σύνταξη και σημασιολογία σύμφωνα με το πρότυπο της RuleML [102], ενώ χρησιμοποιήθηκε ένα δίκτυο οντολογιών πολυμέσων αποτελούμενο από μία οντολογία πυρήνα, οντολογίες θεματικών πεδίων και μία οντολογία για το πρότυπο MPEG-7. Η μηχανή *mandarax* [100] επεξεργάζεται τους κανόνες με σκοπό να εξάγει νέα γνώση. Αυτή η προσέγγιση αντιμετωπίζει το πρόβλημα σε περιπτώσεις στενών θεματικών πεδίων, όταν το λεξιλόγιο εννοιών είναι πολύ μικρό, ενώ μειονέκτημα είναι η μερική ασυμβατότητα των κανόνων με τις οντολογίες που οδηγεί σε μειωμένη εκφραστικότητα και περιορισμένες υπηρεσίες συλλογιστικής. Μία άλλη προσέγγιση στην αναγνώριση σημασιολογικών αντικειμένων σε ακολουθίες βίντεο με χρήση οντολογιών πολυμέσων προτάθηκε στην εργασία [31]. Οι κανόνες στη γλώσσα λογικής πλαισίου (F-logic) καθορίζουν τα εργαλεία ανάλυσης πολυμέσων που θα χρησιμοποιηθούν, ανάλογα με τις έννοιες, τις ιδιότητες και τα χαμηλού επιπέδου χαρακτηριστικά τους, για τον εντοπισμό σημασιολογικών εννοιών ορισμένων στην οντολογία. Τέλος, κανόνες, οντολογίες και εργαλεία ανάλυσης πολυμέσων, αποτελούν τα δεδομένα εισόδου της μηχανής συλλογιστικής KAON [91], ο οποίος αναλαμβάνει να συμπληρώσει αυτόματα μία οντολογία πολυμέσων.

5.7 Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάσαμε ένα δομημένο πλαίσιο γνώσης καθώς και την απαραίτητη μεθοδολογία που πρέπει να ακολουθηθεί για να επιτευχθεί η επιδιωκόμενη σημασιολογική κατάτμηση εικόνων και αναγνώρισης αντικειμένων. Ως γλώσσα αναπαράστασης γνώσης επιλέξαμε τις περιγραφικές λογικές και επιπλέον χρησιμοποιήσαμε τις ασαφείς επεκτάσεις τους για να μπορέσουμε να εντάξουμε στο σύστημά μας το βαθμό ασάφειας που υπεισέρχεται αναπόφευκτα στο στάδιο της ανάλυσης της εικόνας. Εφαρμόσαμε τις υπηρεσίες ασαφούς συλλογιστικής απευθείας στο πρόβλημα της κατάτμησης εικόνων, ως μέθοδο υπολογισμού σημασιολογικής απόστασης δύο περιοχών. Θέλοντας να αποφύγουμε το διαχωρισμό του προβλήματος της κατάτμησης και της αναγνώρισης, των προβλημάτων σε “χαμηλού” και “υψηλού” επιπέδου, εισάγαμε την έννοια της οπτικής λογικής (visual logics) για την ενιαία αντιμετώπιση αυτών των προβλημάτων. Η δομημένη αρχιτεκτονική της οπτικής λογικής βασίζεται στον ορισμό του σώματος ορολογίας (*TBox*), στο σώμα ρόλων (*ABox*) και στο σώμα ρόλων (*RBox*). Το *TBox* αποτελείται από: (α) τις έννοιες-διεγέρτες (trigger concepts), οι οποίες ουσιαστικά αποτελούν τις πρώτες ενδείξεις ότι μία περιοχή οπτικοποιεί μία έννοια, (β) τα οπτικά αξιώματα και τα δομικά αξιώματα, τα οποία χρησιμοποιούνται για να συνθέσουν πιο σύνθετες έννοιες. Η απεικόνιση μίας περιοχής από τους οπτικούς περιγραφείς που τη χαρακτηρίζουν, σε μία έννοια, συνθέτει το σώμα των αξιωμάτων *ABox* στην ερμηνεία του ανοιχτού κόσμου των περιγραφικών λογικών. Ένα κύριο πλεονέκτημα της χρήσης ασαφούς συλλογιστικής είναι η υπηρεσία μέγιστου ελάχιστου φράγματος *glb*, η οποία υπολογίζει το βαθμό συμμετοχής μίας περιοχής στην οπτικοποίηση μίας έννοιας.



Κεφάλαιο 6

Επίλογος

6.1 Συμπεράσματα

Το αντικείμενο έρευνας της διατριβής αυτής ήταν η ενσωμάτωση τεχνολογιών γνώσης στην ανάλυση εικόνων και πιο συγκεκριμένα η χρήση διαφορετικών αναπαραστάσεων γνώσης σε προβλήματα όπως η κατάτμηση εικόνων και βίντεο, η αναγνώριση αντικειμένων, η σημασιολογική περιγραφή και ανάκτηση της οπτικοακουστικής πληροφορίας. Τα ερευνητικά αυτά προβλήματα είναι εγγενώς ασθενώς ορισμένα, καθώς δεν υφίσταται μία μοναδική, αντικειμενικά και καθολικά αποδεκτή λύση. Προκειμένου να καταστεί δυνατή η λύση αυτών των προβλημάτων, είναι απαραίτητο να επαναδιατυπωθούν και να γίνουν ορισμένες υποθέσεις που θα περιορίσουν το πλήθος των αποδεκτών λύσεων. Επιπλέον, παρά τη μεγάλη ερευνητική προσπάθεια, η στάθμη της τεχνολογίας δεν έχει μπορέσει να φτάσει το επίπεδο της ανθρώπινης οπτικής αντίληψης και ο κυριότερος λόγος για αυτό είναι η τεράστια πρότερη γνώση που διαθέτει ο άνθρωπος και ο τρόπος που τη διαχειρίζεται. Η σύμπραξη των τεχνολογιών γνώσης με τεχνικές όρασης υπολογιστών στοχεύει σε αυτό ακριβώς, στην αναπαραστάση και αξιοποίηση της πρότερης γνώσης σε αλγορίθμους σημασιολογικής ανάλυσης εικόνων και βίντεο.

Πιο συγκεκριμένα, στη διατριβή αυτή εξετάσαμε τα διαθέσιμα μοντέλα αναπαραστάσης γνώσης και επιλέξαμε δύο τα οποία ικανοποιούν τις απαιτήσεις που θέσαμε εξ αρχής. Η πρώτη είναι μία υβριδική αναπαραστάση εικόνων και βίντεο με χρήση γράφων και ασαφών συνόλων που χρησιμοποιείται για την σημασιολογική κατάτμηση και αναγνώριση αντικειμένων από το περιεχόμενο. Η ασαφής συνολοθεωρία επιλέχθηκε ως μέσο αναπαραστάσης της αβεβαιότητας, η οποία είναι έμφυτη στα προβλήματα της όρασης υπολογιστών και ανάλυσης εικόνων και βίντεο. Η αβεβαιότητα στη διατριβή αυτή εντοπίζεται στον συσχετισμό των οπτικών χαρακτηριστικών 2Δ ή 3Δ περιοχών της εικόνας ή του βίντεο με σημασιολογικές οντότητες και τα ασαφή σύνολα δίνουν τη δυνατότητα να δοθούν βαθμοί αληθοφάνειας στην ορθότητα ενός τέτοιου συσχετισμού. Από την άλλη, για την εφαρμογή ενός αλγορίθμου επεξεργασίας και ανάλυσης μίας εικόνας ή ενός βίντεο, είναι απαραίτητη η κατάλληλη αναπαραστάσή της σε μία δομή δεδομένων. Καταλληλότερη θεωρήθηκε η δομή γράφων, καθώς η σχετική θεωρία μελετά τις σχέσεις αντικειμένων τα οποία έχουν κάποια συνάφεια μεταξύ τους ή συνδέονται με κάποιον ορισμένο τρόπο, με προφανή την αντιστοιχία με τα αντικείμενα και τις σχέσεις τους μέσα σε μία εικόνα ή ένα βίντεο.

Για την κατανόηση περιεχομένου και τη σημασιολογική περιγραφή αυτού, η βασική απαίτηση ήταν για μία τυπική και εκφραστική γλώσσα αναπαραστάσης γνώσης, η οποία να παρέχει τη δυνατότητα εξαγωγής συμπερασμάτων για τον εμπλουτισμό

της συμβολικής περιγραφής. Για την αναπαράσταση αυτής της γνώσης υιοθετήσαμε το φορμαλισμό των περιγραφικών λογικών, οι οποίες παρέχουν τη δυνατότητα για έλεγχο συνέπειας και υπηρεσίες συλλογιστικής. Επιπλέον οι περιγραφικές λογικές αποτελούν τη βάση της γλώσσας OWL που είναι η πρότυπη γλώσσα της W3C για την αναπαράσταση οντολογιών στον σημασιολογικό ιστό.

Δεδομένης της πρότερης γνώσης και της αναπαράστασής αυτής, τη μετάβαση από τα αριθμητικά χαρακτηριστικά του σήματος στα συμβολικά-εννοιολογικά, τη διατυπώσαμε ως ταίριασμα περιοχών της εικόνας με πρότυπα στιγμιότυπα εννοιών της οντολογίας θεματικού πεδίου. Πιο συγκεκριμένα, συνδέσαμε τις έννοιες της οντολογίας θεματικού πεδίου με πρότυπα στιγμιότυπα οπτικών περιγραφών μέσω της οντολογίας οπτικών περιγραφών. Η ανάκτηση αυτών των στιγμιότυπων και το ταίριασμα των οπτικών περιγραφών με τους αντίστοιχους που εξάγονται από την υπό ανάλυση εικόνα ή βίντεο, αποτέλεσε ένα σημαντικό κομμάτι της αλληλεπίδρασης των αλγορίθμων ανάλυσης με τις τεχνολογίες γνώσεις.

Τα χαρακτηριστικά - οπτικά και συμβολικά - που έχουν εξαχθεί στο προηγούμενο στάδιο της ανάλυσης και έχουν αποθηκευτεί στη δομή του σχεσιακού γράφου χαρακτηριστικών, αξιοποιήθηκαν στα πλαίσια της σημασιολογικής κατάτμησης εικόνων και βίντεο. Αντιμετωπίσαμε το πρόβλημα της από κοινού κατάτμησης και αναγνώρισης αντικειμένων, στο πλαίσιο της συγχώνευσης γειτονικών περιοχών, οι οποίες έχουν μία κοινή σημασιολογική ερμηνεία. Σε αντιδιαστολή με την συνήθη πρακτική της σειριακής εφαρμογής των διαδικασιών της κατάτμησης και της αναγνώρισης, ενσωματώσαμε στον αλγόριθμο κατάτμησης χαρακτηριστικά που παράχθηκαν κατά τη διαδικασία της αναγνώρισης, και εφαρμόσαμε υβριδικά (οπτικά και συμβολικά) κριτήρια για την συγχώνευση περιοχών. Ο κύριος στόχος της σημασιολογικής κατάτμησης ήταν να επιτύχει καλύτερη κατάτμηση αλλά να ανατροφοδοτήσει με βελτιωμένα στοιχεία τη διεργασία της αναγνώρισης, χαρακτηρίζοντας τις επιμέρους περιοχές που συνθέτουν την εικόνα ή το βίντεο.

Τέλος, η σημασιολογική ανάλυση εικόνων και βίντεο ολοκληρώνεται με την εξαγωγή μεταδεδομένων, τα οποία περιγράφουν το περιεχόμενο σε ένα εννοιολογικό επίπεδο, ισοδύναμο με αυτό που αντιλαμβάνεται και επιθυμεί να διαχειριστεί ένας άνθρωπος. Χρησιμοποιήσαμε το τυπικό, δομημένο πλαίσιο γνώσης των ασαφών περιγραφικών λογικών, για τον έλεγχο της ορθότητας και τον εμπλουτισμό της εξαχθείσας σημασιολογικής πληροφορίας. Αναδείξαμε τη χρησιμότητα του αυστηρού πλαισίου γνώσης, μέσω της δυνατότητας που έχει για εύκολη επέκταση, διαμοιρασμό και επαναχρησιμοποίηση. Επιπρόσθετα, η παραγόμενη σημασιολογική περιγραφή του περιεχομένου, είναι δομική, καθώς ο χαρακτηρισμός αφορά τόσο όλης την εικόνα, όσο και τα αντικείμενα που περιέχονται σε αυτήν.

6.2 Συνεισφορά και Δημοσιεύσεις

Υβριδική αναπαράσταση εικόνων και βίντεο με γράφους και ασαφή σύνολα
Αυτονόητο προαπαιτούμενο προκειμένου να ενσωματώσουμε τεχνολογίες γνώσεις σε προβλήματα ανάλυσης εικόνας, είναι να ορίσουμε τι είδους γνώση χρειαζόμαστε και πώς μπορούμε να την αναπαραστήσουμε πιο κατάλληλα. Ουσιαστικά χρειαζόμαστε κάποιου είδους αναπαράσταση πρότερης γνώσης για να δώσουμε εννοιολογική περιγραφή στα οπτικά χαρακτηριστικά που εξάγουμε απευθείας από την εικόνα. Παράλληλα, η αβεβαιότητα αφθονεί στα προβλήματα της όρασης υπολογιστών, καθιστώντας χρήσιμο εργαλείο την ασαφή συνολοθεωρία για την αναπαράσταση αυτού του είδους

πληροφορίας. Έτσι λοιπόν, προτείνουμε τη χρησιμοποίηση των ασαφών συνόλων για την αναπαράσταση των πιθανών εννοιών που σχετίζονται (με κάποιο βαθμό συμμετοχής) με μία εικόνα και τις επιμέρους περιοχές που την αποτελούν.

Επιπρόσθετα, χρειαζόμαστε μία δομή που να μπορεί να διαχειρίζεται αποδοτικά τα οπτικά χαρακτηριστικά και τη χωρική πληροφορία που εξάγεται από μία εικόνα. Για αυτόν τον σκοπό επιλέξαμε τον σχεσιακό γράφο χαρακτηριστικών. Η θεωρία γράφων αποτελεί μία καλά μελετημένη επιστημονική περιοχή που μας παρέχει τα κατάλληλα εφόδια για να προβλήματα ανάλυσης εικόνας. Πιο συγκεκριμένα, οι κόμβοι του γράφου αντιστοιχούν σε μη-επικαλυπτόμενες περιοχές οι οποίες περιγράφονται με κάποια χαρακτηριστικά.

Η καινοτομία της προτεινόμενης προσέγγισης είναι η ενσωμάτωση του ασαφούς συνόλου εννοιών, που σχετίζονται με κάθε περιοχή, στα χαρακτηριστικά των αντίστοιχων κόμβων του γράφου αναπαράστασης της εικόνας. Οι σταθμισμένες ακμές του γράφου αντιστοιχούν στις χωρικές σχέσεις μεταξύ των επιμέρους περιοχών και το βάρος τους υποδηλώνει το βαθμό σημασιολογικής ομοιότητας αυτών των δύο συνδεόμενων περιοχών. Επιπλέον, προτείναμε την επέκταση του σχεσιακού γράφου χαρακτηριστικών με χρονικές ακμές για την αναπαράσταση ακολουθιών βίντεο. Οι δημοσιεύσεις που σχετίζονται με το συγκεκριμένο κομμάτι της διατριβής είναι δύο άρθρα σε πρακτικά συνεδρίων [6, 7] και αποσπάσματα σε δύο άρθρα περιοδικών [9, 86].

Σημασιολογική κατάτμηση εικόνων και βίντεο

Η πρώτη εφαρμογή της ενσωμάτωσης γνώσης σε προβλήματα ανάλυσης εικόνας ήταν η σημασιολογική κατάτμηση. Κύριος στόχος της σημασιολογικής κατάτμησης είναι να επιτύχει καλύτερη κατάτμηση της εικόνας και να χαρακτηρίσει και να κατατάξει τις επιμέρους περιοχές που την συνθέτουν. Η καινοτομία της προτεινόμενης μεθοδολογίας έγκειται στην τροποποίηση καθιερωμένων αλγορίθμων κατάτμησης επέκτασης περιοχών, ώστε να ενσωματώσουν χαρακτηριστικά ενδιάμεσου επιπέδου. Σε αντίθεση με μία πλειάδα τεχνικών που τοποθετούν σειριακά τους αλγορίθμους κατάτμησης εικόνας και αναγνώρισης αντικειμένων, προτείνουμε την ενσωμάτωση της αναγνώρισης μέσα στον αλγόριθμο της κατάτμησης. Τα χαρακτηριστικά ενδιάμεσου επιπέδου είναι σημασιολογικές οντότητες (ή αλλιώς σημασιολογικές ετικέτες) οι οποίες αντιστοιχούν σε κάθε περιοχή και αναπαριστώνται με ασαφή σύνολα. Τα κριτήρια συγχώνευσης δύο περιοχών πλέον δεν βασίζονται αποκλειστικά στα οπτικά χαρακτηριστικά τους (οπτική ομοιότητα), αλλά και στα σημασιολογικά χαρακτηριστικά όπως τα αποδίδουν τα δύο ασαφή σύνολα εννοιών (σημασιολογική ομοιότητα).

Τροποποιήσαμε δύο γνωστούς αλγορίθμους επέκτασης περιοχών, το μετασχηματισμό watershed και τον αλγόριθμο RSST, ώστε να χρησιμοποιούν τα σημασιολογικά χαρακτηριστικά ως κριτήρια συγχώνευσης δυο περιοχών και τερματισμού τους. Χρησιμοποιήσαμε τελεστές ασαφών συνόλων (τ-νόρμες, τ-κονόρμες και συνάθροισης) για τον υπολογισμό της σημασιολογικής ομοιότητας δύο περιοχών και την εξαγωγή των νέων ετικετών μετά την συγχώνευσή. Επιπρόσθετα, επεκτείναμε τη μεθοδολογία σημασιολογικής κατάτμησης σε ακολουθίες βίντεο, επιτυγχάνοντας χωροχρονική εξαγωγή, εντοπισμό και παρακολούθηση αντικειμένων, με ταυτόχρονη μείωση της υπολογιστικής πολυπλοκότητας που έχει η χωροχρονική ανάλυση βίντεο.

Οι δημοσιεύσεις που σχετίζονται με το συγκεκριμένο κομμάτι της διατριβής είναι δύο άρθρα σε πρακτικά συνεδρίων [8, 115] για την περίπτωση της κατάτμησης εικόνων και ένα [40] για την περίπτωση του βίντεο. Επιπλέον, η συνεισφορά περιγράφεται διεξοδικότερα σε δύο άρθρα περιοδικών [9, 114] και σε ένα κεφάλαιο βιβλίου [130].

Οπτική λογική

Μία βασική ιδιότητα των τυπικών γλωσσών αναπαράστασης γνώσης, όπως για παράδειγμα η OWL, είναι η δυνατότητα χρήσης υπηρεσιών συλλογιστικής για την τον έλεγχο της ορθότητας και την επέκταση της γνώσης. Παρουσιάσαμε ένα δομημένο και ιεραρχικό πλαίσιο αναπαράστασης γνώσης στηριζόμενο στις περιγραφικές λογικές και τις ασαφείς επεκτάσεις τους για να μπορούμε να εντάξουμε στο σύστημά μας το βαθμό ασάφειας που υπεισέρχεται αναπόφευκτα στο στάδιο της ανάλυσης. Αναδείξαμε την σημασιολογία των εννοιών και των ρόλων των περιγραφικών λογικών σε σχέση με το πρόβλημα της ανάλυσης εικόνας και περιγράψαμε τη δυνατότητα εύκολης επέκτασης και επαναχρησιμοποίησης της βάσης γνώσης σε ένα νέο σύστημα.

Το εν λόγω πλαίσιο το ονομάσαμε οπτική λογική, καθώς στηρίζεται σε μετρήσιμα οπτικά χαρακτηριστικά πάνω στα οποία θεμελιώνεται ο ορισμός πιο σύνθετων εννοιών χρησιμοποιώντας οπτικά αξιώματα. Οι έννοιες-διεγέρτες είναι έννοιες που μπορούν να ανιχνευθούν απευθείας με ορισμένους (απλούς) υπολογισμούς πάνω στην εικόνα και εν συνεχεία “διεγείρουν” τη διαδικασία επέκτασης της γνώσης. Η διαδικασία αυτή βασίζεται σε δύο ειδών αξιώματα, τα απλά και τα δομικά. Τα πρώτα ορίζουν νέες έννοιες χρησιμοποιώντας μόνο εκφράσεις εννοιών-διεγερτών, ενώ τα δεύτερα μπορούν να χρησιμοποιήσουν και ρόλους για να εκφράσουν πιο σύνθετες έννοιες.

Οι δημοσιεύσεις που σχετίζονται με το συγκεκριμένο κομμάτι της διατριβής είναι ένα άρθρο πάνω στη συλλογιστική πολυμέσων [113] και άλλα δύο σχετικά με τη χρήση ασαφών περιγραφικών λογικών για ανάλυση και ανάκτηση πολυμέσων, σε συνέδριο [117] και σε περιοδικό [116].

Κατανόηση περιεχομένου βάσει υπηρεσιών (ασαφούς) συλλογιστικής

Στην κατανόηση του περιεχομένου μιας εικόνας, που αποτελεί ένα από τα σημαντικά προβλήματα, συνεισφέρει η αρχιτεκτονική της οπτικής λογικής με τη χρήση υπηρεσιών συλλογιστικής. Ονομάσαμε οπτική συλλογιστική το σύνολο των υπηρεσιών συλλογιστικής που προσφέρουν οι ασαφείς περιγραφικές λογικές, οι οποίες υπηρεσίες χρησιμοποιούνται έμπρακτα σε προβλήματα ανάλυσης εικόνας. Πιο συγκεκριμένα, προτείναμε τη χρήση του μέγιστου ελάχιστου φράγματος για τον υπολογισμό της σημασιολογικής ομοιότητας δύο περιοχών ή εικόνων, όπως αυτός συνεπάγεται από τη βάση γνώσης. Η βάση γνώσης αποτελείται αφενός από τα αξιώματα (σώμα ορολογίας - TBox) και αφετέρου από τη συσχέτιση των εννοιών που προκύπτει από τη διαδικασία εντοπισμού και αναγνώρισης (όπως π.χ. στην περίπτωση της σημασιολογικής κατάτμησης).

Οι δημοσιεύσεις που σχετίζονται με το συγκεκριμένο κομμάτι της διατριβής είναι δύο άρθρα σε πρακτικά συνεδρίων [113, 10] και ένα άρθρο σε περιοδικό [116].

6.3 Μελλοντικές Επεκτάσεις

Είναι δεδομένο ότι η ερευνητική δραστηριότητα είναι αέναη και σε καμία περίπτωση δεν ολοκληρώνεται στο χρονικό ορίζοντα μίας διδακτορικής διατριβής. Όσο περισσότερο εμβαθύνει κανείς σε ένα αντικείμενο, τόσο περισσότερο δημιουργούνται νέες απορίες και προβλήματα αλλά και ιδέες για επίλυση αυτών. Με βάση την παρούσα διατριβή προτείνουμε ορισμένες κατευθύνσεις προς τις οποίες μπορεί να επεκταθεί η έρευνα στο ευρύτερο πλαίσιο της σημασιολογικής ανάλυσης πολυμέσων και της κατανόησης

περιεχομένου. Συνοπτικά προτείνονται:

1. Χρήση γεωμετρίας και τμηματικών μοντέλων για τον εντοπισμό και την αναγνώριση σύνθετων αντικειμένων. Αφορμή αποτελούν οι εργασίες [21, 62] οι οποίες μπορούν να επεκταθούν με χρήση μερονομικών (partonomic) σχέσεων και διατύπωση του προβλήματος ως ταίριασμα υπογράφων.
2. Ενσωμάτωση πιθανοτικών μοντέλων στο σχεσιακό γράφο χαρακτηριστικών για την καλύτερη αναπαράσταση των δεδομένων. Ενδεικτικά, σε παρεμφερείς εργασίες έχει προταθεί η χρήση εξαρτημένων τυχαίων πεδίων (CRFs) [111], μαρκοβιανών τυχαίων πεδίων (MRFs) [79] και ο αλγόριθμος μεγιστοποίησης προσδοκίας (EM) [59].
3. Επέκταση της αρχιτεκτονικής της οπτικής λογικής με προσθήκη επιπλέον εννοιών, αξιωμάτων ή τεχνικών εξαγωγής εννοιών διεγερτών. Επίσης προτείνεται η περαιτέρω αξιοποίηση των περιγραφικών λογικών για τη σημασιολογική ερμηνεία [96] και ανάκτηση πολυμεσικού περιεχομένου με χρήση σημασιολογικών ερωτημάτων [81].

Πιο αναλυτικά, η χρήση πρότερης γνώσης για τα μέρη που συνθέτουν ορισμένα αντικείμενα και η γεωμετρία που τα διέπει, μπορεί να βοηθήσει στη δημιουργία υποθέσεων για την ύπαρξη του αντικειμένου αν εντοπιστούν ορισμένες από τις συνιστώσες του. Για παράδειγμα με δεδομένη τη γεωμετρία του προσώπου, ο εντοπισμός ενός ματιού μπορεί να βοηθήσει στην κατάτμηση και την αναγνώριση και άλλων μερών του προσώπου, όπως η μύτη και το στόμα. Έτσι μπορεί να αντιμετωπιστεί πιο αποτελεσματικά το πρόβλημα της επικάλυψης ή κακής αρχικής κατάτμησης. Μία πιθανή μοντελοποίηση των μερών με γράφο, μετατρέπει το πρόβλημα σε ταίριασμα υπογράφων· μεταξύ του γράφου αναπαράστασης της εικόνας και του μοντέλου του αντικειμένου. Η εύρεση της βέλτιστης λύσης στο πρόβλημα του ταυρίσματος υπογράφων είναι ισοδύναμη με το πρόβλημα της εύρεσης όλων των ισομορφισμών των όλων των υπογράφων και είναι NP πολυπλοκότητας. Για αυτό έχουν προταθεί στοχαστικές τεχνικές, προσεγγιστικές (ιεραρχική αποδόμηση σε υπογράφους), μη ακριβές (ή ισοδύναμα, ανεκτικό σε λάθη) ταίριασμα [65], μετρικές ομοιότητας γράφων και ελαχιστοποίηση συναρτήσεων κόστους με τομές γράφων (graph cuts) [23].

Στη διατριβή αυτή χρησιμοποιήσαμε την ασαφή συνολοθεωρία για την αναπαράσταση της αβεβαιότητας στη διαδικασία της αναγνώρισης εννοιών. Η ερμηνεία των περιοχών της εικόνας ή του βίντεο μπορεί να μοντελοποιηθεί με πιο σύνθετα πιθανοτικά μοντέλα όπως είναι τα CRFs και τα MRFs. Με ένα τέτοιο μοντέλο μπορούμε να κάνουμε υποθέσεις βάσει των οπτικών χαρακτηριστικών αλλά και των υποθέσεων για τις γειτονιές περιοχές, δίνοντας μία νέα ερμηνεία στο οπτικό εννοιολογικό πλαίσιο. Το αρνητικό αυτής της πρακτικής είναι η σαφής διόγκωση της υπολογιστικής πολυπλοκότητας σε σχέση με αυτή της αναπαράστασης με τα ασαφή σύνολα.

Τέλος, στηριζόμενοι στις ασαφείς περιγραφικές λογικές, χρησιμοποιήσαμε μηχανισμούς συλλογιστικής για να δεικτοδοτήσουμε και να αποθηκεύσουμε την εξαγόμενη σημασιολογική περιγραφή σε μία βάση γνώσης. Αυτό που μπορεί να διερευνηθεί είναι η δυνατότητα εκτέλεσης ερωτημάτων σε μία τέτοια βάση γνώσης, τα οποία με χρήση των κατάλληλων υπηρεσιών συλλογιστικής μπορούν να λειτουργήσουν ως μηχανισμός σημασιολογικής ανάκτησης και κατάταξης εικόνων. Επιπλέον, η προτεινόμενη προσέγγιση της οπτικής λογικής παρέχει μία αρθρωτή αρχιτεκτονική που δίνει δυνατότητα για επέκταση είτε με προσθήκη νέων εννοιών, είτε νέων αξιωμάτων, είτε ακόμα και

νέων τεχνικών εξαγωγής εννοιών διεγερτών (π.χ. ένας αλγόριθμος εντοπισμού προσώπων). Η προσθήκη νέων στοιχείων και μεθοδολογιών εξαγωγής διεγερτών μπορεί να εμπλουτίσει την αρχική γνώση και να δώσει μεγαλύτερη ευρωστία στο σύστημα μας.

□

Παράρτημα Α

Κατάλογος Απόδοσης Όρων

A.1 Κατάλογος Αγγλικών Όρων

a priori knowledge	πρότερη γνώση
adjacency matrix	πίνακας γειτνίασης
aggregation	συνάθροιση
aggregation operator	τελεστής συνάθροισης
angular radial transformation (ART)	γωνιακός ακτινικός μετασχηματισμός
assertion box (ABox)	σώμα ισχυρισμών
at-least	περιορισμός το-λιγότερο
at-most	περιορισμός το-πολύ
atomic concept	ατομική έννοια
atomic role	ατομικός ρόλος
attributed relational graph (ARG)	σχεσιακός γράφος χαρακτηριστικών
back-projection	πίσω-προβολή
bayesian network	δίκτυο bayes
block of frames (BoF)	μπλοκ από καρτέ
bottom concept	κενή έννοια
bottom-up (approach)	κάτωθεν/εκ των κάτω (προσέγγιση)
bounding box	περικλείον πλαίσιο
browsing	πλοήγηση
cardinality	πληθυστικότητα
catchment basin	κοιλιάδα περισυλλογής
classification	ταξινόμηση
classifier	ταξινομητής
clustering	συσταδοποίηση
coarseness	τραχύτητα
combinatorial optimization	συνδυαστική βελτιστοποίηση
concept assertions	ισχυρισμοί εννοιών
concept	έννοια
conditional random fields	εξαρτημένα τυχαία πεδία

conjunction	τομή/σύζευξη
context	εννοιολογικό πλαίσιο
contour-based shape descriptor	περιγραφέας σχήματος με βάση το περίγραμμα
core ontology	οντολογία πυρήνα
crisp set	κλασικό σύνολο
decision trees	δένδρα αποφάσεων
decidable	αποφασίσιμος
decomposition	αποδόμηση
degree of membership	βαθμός συμμετοχής
description logics (DLs)	περιγραφικές λογικές (ΠΛ)
directionality	κατεύθυνση/προσανατολισμός
disjunction	ένωση/διάζευξη
domain knowledge	γνώση ειδικού πεδίου
dominant color descriptor	περιγραφέας κύριων/χαρακτηριστικών χρωμάτων
dominant concept	κύρια έννοια
edge (of a graph)	ακμή (γράφου)
efficiency	απόδοση
equivalence	ισοδυναμία
exists restriction	υπαρξιακός περιορισμός
expectation maximization (EM)	μεγιστοποίηση προσδοκίας
explicit	ρητός
extensional knowledge	ρητή γνώση
first order logic	λογικής πρώτης τάξης
flooding	πλημμυρισμός
formalism	φορμαλισμός
fragment identifier	αναγνωριστικό τμήματος
frame	καρέ
frame logic (f-logic)	λογική πλαισίου
fusion	σύντηξη
gaussian mixture model (GMM)	μείγμα μοντέλων γκαουσιανών κατανο- μών
general negation	γενικευμένη άρνηση
gradient	παράγωγος
graph cuts	τομές γραφών
greatest lower bound (GLB)	μέγιστο ελάχιστο φράγμα
ground truth	σύνολο δεδομένων αλήθειας
hardware	υλισμικό
height (of a fuzzy set)	ύψος (ενός ασαφούς συνόλου)
homogeneous texture descriptor	περιγραφέας ομοιογενούς υφής

ill-posed	ελλιπώς/ασθενώς ορισμένος
implication	συνεπαγωγή
implicit	υπονοούμενος
indexing	δεικτοδότηση
individual	άτομο
induced subgraph	επαγόμενος υπογράφος
instance relation	σχέση στιγμιότυπου
intentional knowledge	υπονοούμενη γνώση
interoperability	διαλειτουργικότητα
inverse role	αντίστροφος ρόλος
invocation	επίκληση
labeling	επισήμανση
marker	σημαδευτής
markov random fields	μαρκοβιανά τυχαία πεδία
membership function	συνάρτηση συμμετοχής
metadata	μεταδεδομένα
minimum spanning tree	ελάχιστο συνδετικό δένδρο
modality	τροπικότητα
multiclass	πολυκλασικός
multimedia structure ontology (MSO)	οντολογία δομής πολυμέσων
multiple cue	πολλαπλά στοιχεία-πρότυπα
namespace	ονοματοχώρος
non-rigid (motion)	σπονδυλωτή (κίνηση)
occlusion	επικάλυψη
open-world assumption	υπόθεση ανοικτού κόσμου
oversegmentation	υπερκατάτμηση
performance	επίδοση
pixel	εικονοστοιχείο
power set	δυναμοσύνολο
primitives	βασικοί τύποι
propagation function	συνάρτηση διάδοσης/διασποράς
propagator	διασπορέας
prototype instance (PI)	πρότυπο παράδειγμα/στιγμιότυπο
reasoning	συλλογιστική
recursive shortest spanning tree (RSST)	αναδρομικά ελάχιστο συνδετικό δένδρο
region growing	επέκταση περιοχών
region merging	συγχώνευση περιοχών
region splitting	διαίρεση περιοχών
region-based shape descriptor	περιγραφέας σχήματος με βάση την περιοχή

regularity	κανονικότητα
resource identifier	αναγνωριστικό πόρου
resources	πόροι
role assertions	ισχυρισμοί ρόλων
role axiom	αξίωμα ρόλου
role box (RBox)	σώμα ρόλων
scalable	κλιμακούμενος
semantic gap	σημασιολογικό κενό
semantic region growing (SRG)	σημασιολογική επέκταση περιοχών
semantic web	σημασιολογικός ιστός
serialization	σειριακοποίηση
sub-concept	υπό-έννοια
sub-role	υπό-ρόλος
support (of a fuzzy set)	στήριγμα (ενός ασαφούς συνόλου)
support vector machines (SVMs)	μηχανές διανυσμάτων στήριξης
taxonomy	ταξινομία
terminology box (TBox)	σώμα ορολογίας
texture browsing descriptor	περιγραφέας αναζήτησης υφής
top concept	καθολική έννοια
top-down (approach)	άνωθεν/καθοδική (προσέγγιση)
tractable	βατός
transitive role	μεταβατικός ρόλος
trigger concepts	έννοιες διεγέρτες
trigger	διεγέρτης
unary predicates	μοναδιαία κατηγορήματα
undersegmentation	υποκατάτμηση
uniform resource identifier (URI)	ομοιόμορφο αναγνωριστικό πόρων
universal set	γενικό σύνολο/σύνολο αναφοράς
value restriction	περιορισμός τιμής
vertex (of a graph)	κόμβος (γράφου)
video shot	πλάνο
visual axiom	οπτικό αξίωμα
visual context	οπτικό εννοιολογικό πλαίσιο
visual descriptor ontology (VDO)	οντολογία οπτικών περιγραφέων
visual logics	οπτική λογική
visual ontology	οπτική οντολογία
visual reasoning	οπτική συλλογιστική
watershed	υδροκρίτης
watershed line	υδατογραμμή
wavelet	κυματίδιο

A.2 Κατάλογος Ελληνικών Όρων

ακμή (γράφου)	edge (of a graph)
αναγνωριστικό πόρου	resource identifier
αναγνωριστικό τμήματος	fragment identifier
αναδρομικά ελάχιστο	recursive shortest
συνδετικό δένδρο	spanning tree (RSST)
αντίστροφος ρόλος	inverse role
άνωθεν/καθοδική (προσέγγιση)	top-down (approach)
αξίωμα ρόλου	role axiom
αποδόμηση	decomposition
απόδοση	efficiency
αποφασίσιμος	decidable
ατομική έννοια	atomic concept
ατομικός ρόλος	atomic role
άτομο	individual
βαθμός συμμετοχής	degree of membership
βασικοί τύποι	primitives
βατός	tractable
γενικευμένη άρνηση	general negation
γενικό σύνολο/σύνολο αναφοράς	universal set
γνώση ειδικού πεδίου	domain knowledge
γωνιακός ακτινικός μετασχηματισμός	angular radial transformation (ART)
δεικτοδότηση	indexing
δένδρα αποφάσεων	decision trees
διαίρεση περιοχών	region splitting
διαλειτουργικότητα	interoperability
διασπορέας	propagator
διεγέρτης	trigger
δίκτυο bayes	bayesian network
δυναμοσύνολο	power set
σύνολο δεδομένων αλήθειας	ground truth
εικονοστοιχείο	pixel
ελάχιστο συνδετικό δένδρο	minimum spanning tree
ελλιπώς/ασθενώς ορισμένος	ill-posed
έννοια	concept
έννοιες διεγέρτες	trigger concepts
εννοιολογικό πλαίσιο	context
ένωση/διάζευξη	disjunction
εξαρτημένα τυχαία πεδία	conditional random fields
επαγόμενος υπογράφος	induced subgraph

επέκταση περιοχών	region growing
επίδοση	performance
επικάλυψη	occlusion
επίκληση	invocation
επισήμανση	labeling
ισοδυναμία	equivalence
ισχυρισμοί εννοιών	concept assertions
ισχυρισμοί ρόλων	role assertions
καθολική έννοια	top concept
κανονικότητα	regularity
καρέ	frame
κατεύθυνση/προσανατολισμός	directionality
κάτωθεν/εκ των κάτω (προσέγγιση)	bottom-up (approach)
κενή έννοια	bottom concept
κλασικό σύνολο	crisp set
κλιμακούμενος	scalable
κοιλάδα περισυλλογής	catchment basin
κόμβος (γράφου)	vertex (of a graph)
κυματίδιο	wavelet
κύρια έννοια	dominant concept
λογική πλαισίου	frame logic (f-logic)
λογικής πρώτης τάξης	first order logic
μαρκοβιανά τυχαία πεδία	markov random fields
μέγιστο ελάχιστο φράγμα	greatest lower bound (GLB)
μεγιστοποίηση προσδοκίας	expectation maximization (EM)
μείγμα μοντέλων γκαουσιανών κατανομών	gaussian mixture model (GMM)
μεταβατικός ρόλος	transitive role
μεταδεδομένα	metadata
μηχανές διανυσμάτων στήριξης	support vector machines (SVMs)
μοναδιαία κατηγορήματα	unary predicates
μπλοκ από καρέ	block of frames (BoF)
ομοιόμορφο αναγνωριστικό πόρων	uniform resource identifier (URI)
ονοματοχώρος	namespace
οντολογία δομής πολυμέσων	multimedia structure ontology (MSO)
οντολογία οπτικών περιγραφών	visual descriptor ontology (VDO)
οντολογία πυρήνα	core ontology
οπτική λογική	visual logics
οπτική οντολογία	visual ontology
οπτική συλλογιστική	visual reasoning
οπτικό αξίωμα	visual axiom

οπτικό εννοιολογικό πλαίσιο	visual context
παράγωγος	gradient
περιγραφέας αναζήτησης υφής	texture browsing descriptor
περιγραφέας κύριων/χαρακτηριστικών χρωμάτων	dominant color descriptor
περιγραφέας ομοιογενούς υφής	homogeneous texture descriptor
περιγραφέας σχήματος με βάση την περιοχή	region-based shape descriptor
περιγραφέας σχήματος με βάση το περίγραμμα	contour-based shape descriptor
περιγραφικές λογικές (ΠΛ)	description logics (DLs)
περικλείον πλαίσιο	bounding box
περιορισμός τιμής	value restriction
περιορισμός το-λιγότερο	at-least
περιορισμός το-πολύ	at-most
πίνακας γειτνίασης	adjacency matrix
πίσω-προβολή	back-projection
πλάνο	video shot
πληθυκότητα	cardinality
πλημμυρισμός	flooding
πλοήγηση	browsing
πολλαπλά στοιχεία-πρότυπα	multiple cue
πολυκλασικός	multiclass
πόροι	resources
πρότερη γνώση	a priori knowledge
πρότυπο παράδειγμα/στιγμιότυπο	prototype instance (PI)
ρητή γνώση	extensional knowledge
ρητός	explicit
σειριακοποίηση	serialization
σημαδευτής	marker
σημασιολογική επέκταση περιοχών	semantic region growing (SRG)
σημασιολογικό κενό	semantic gap
σημασιολογικός ιστός	semantic web
σπονδυλωτή (κίνηση)	non-rigid (motion)
στήριγμα (ενός ασαφούς συνόλου)	support (of a fuzzy set)
συγχώνευση περιοχών	region merging
συλλογιστική	reasoning
συνάθροιση	aggregation
συνάρτηση διάδοσης/διασποράς	propagation function
συνάρτηση συμμετοχής	membership function

συνδυαστική βελτιστοποίηση	combinatorial optimization
συνεπαγωγή	implication
σύντηξη	fusion
συσταδοποίηση	clustering
σχέση στιγμιότυπου	instance relation
σχεσιακός γράφος χαρακτηριστικών	attributed relational graph (ARG)
σώμα ισχυρισμών	assertion box (ABox)
σώμα ορολογίας	terminology box (TBox)
σώμα ρόλων	role box (RBox)
ταξινόμηση	classification
ταξινομητής	classifier
ταξινομία	taxonomy
τελεστής συνάθροισης	aggregation operator
τομές γράφων	graph cuts
τομή/σύζευξη	conjunction
τραχύτητα	coarseness
τροπικότητα	modality
υδατογραμμή	watershed line
υδροκρίτης	watershed
υλισμικό	hardware
υπαρξιακός περιορισμός	exists restriction
υπερκατάτμηση	oversegmentation
υπό-έννοια	sub-concept
υπόθεση ανοικτού κόσμου	open-world assumption
υποκατάτμηση	undersegmentation
υπονοούμενη γνώση	intentional knowledge
υπονοούμενος	implicit
υπό-ρόλος	sub-role
ύψος (ενός ασαφούς συνόλου)	height (of a fuzzy set)
φορμαλισμός	formalism

A.3 Ακρωνύμια

ABox	assertion box	σώμα ισχυρισμών
ARG	attributed relational graph	σχεσιακός γράφος χαρακτηριστικών
ART	angular radial transformation	γωνιακός ακτινικός μετασχηματισμός
BoF	block of frames	μπλοκ από καρέ
CRFs	conditional random fields	εξαρτημένα τυχαία πεδία
DLs	description logics	περιγραφικές λογικές
EM	expectation maximization	μεγιστοποίηση προσδοκίας
GLB	greatest lower bound	μέγιστο ελάχιστο φράγμα
GMM	gaussian mixture model	μείγμα μοντέλων γκαουσιανών κατανομών
KAA	knowledge assisted analysis	ανάλυση με χρήση γνώσης
MRFs	markov random fields	μαρκοβιανά τυχαία πεδία
MSO	multimedia structure ontology	οντολογία δομής πολυμέσων
PI	prototype instance	πρότυπο παράδειγμα/στιγμιότυπο
RBox	role box	σώμα ρόλων
RSST	recursive shortest spanning tree	αναδρομικά ελάχιστο συνδετικό δένδρο
SRG	semantic region growing	σημασιολογική επέκταση περιοχών
SRSST	semantic recursive shortest spanning tree	σημασιολογικό αναδρομικά ελάχιστο συνδετικό δένδρο
SW	semantic watershed	σημασιολογικό watershed
SVMs	support vector machines	μηχανές διανυσμάτων στήριξης
TBox	terminology box	σώμα ορολογίας
URI	uniform resource identifier	ομοίμορφο αναγνωριστικό πόρων
VDO	visual descriptor ontology	οντολογία οπτικών περιγραφών
ανν	if and only if	αν και μόνο αν
ΠΑ	description logics	περιγραφικές λογικές

□

Παράρτημα Β

Κατάλογος δημοσιεύσεων

Β.1 Περιοδικά

1. N. Simou, Th. Athanasiadis, S. Kollias, G. Stamou. Semantic Adaptation of Neural Network Classifiers in Image Segmentation. In Neural Network world, Institute of Computer Science ASCR, Volume 19, Number 5, pp. 561-579, October 2009.
2. N. Simou, Th. Athanasiadis, G. Stoilos, S. Kollias. Image Indexing and Retrieval using Expressive Fuzzy Description Logics. In Signal, Image and Video Processing, Springer, Volume 2, Number 4, pp. 321-335, December 2008.
3. Ph. Mylonas, Th. Athanasiadis, M. Wallace, Y. Avrithis, S. Kollias. Semantic Representation of Multimedia Content - Knowledge Representation and Semantic Indexing. In Multimedia Tools and Applications, Springer, Volume 39, Number 3, pp. 293-327, September 2008.
4. Th. Athanasiadis, Ph. Mylonas, Y. Avrithis, S. Kollias. Semantic Image Segmentation and Object Labeling. In Transactions on Circuits and Systems for Video Technology, IEEE, Volume 17, Number, pp. 298-312, March 2007.
5. M. Wallace, Th. Athanasiadis, Y. Avrithis, A. Delopoulos, S. Kollias. Integrating Multimedia Archives: The Architecture and the Content Layer. In Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE, Volume 36, Number 1, pp. 34-52, January 2006.
6. Th. Athanasiadis, M. Wallace, K. Karpouzis, S. Kollias. Utilization of Evidence Theory in the Detection of Salient Regions in Successive CT Images. In Oncology Reports, Spandidos Publications, Volume 15, Number 4, pp. 1071-1076, April 2006.

B.2 Κεφάλαια Βιβλίων

1. Th. Athanasiadis, Ph. Mylonas, G. Papadopoulos, V. Mezaris, Y. Avrithis, I. Kompatsiaris, M. Strintzis. Knowledge Driven Segmentation and Classification. In *Multimedia Semantics: Metadata, Analysis and Interaction*, Willey, in press.

B.3 Συνέδρια

1. Th. Athanasiadis, N. Simou, G. Papadopoulos, R. Benmokhtar, K. Chandramouli, V. Tzouvaras, V. Mezaris, M. Phiniketos, Y. Avrithis, Y. Kompatsiaris, B. Huet, E. Izquierdo. Integrating Image Segmentation and Classification for Fuzzy Knowledge-based Multimedia Indexing. In *Proceedings of 15th International MultiMedia Modeling Conference (MMM 2009)*, January 2009, Sophia Antipolis, France.
2. G. Papadopoulos, Th. Athanasiadis, N. Simou, R. Benmokhtar, K. Chandramouli, V. Tzouvaras, V. Mezaris, M. Phiniketos, Y. Avrithis, I. Kompatsiaris, B. Huet, E. Izquierdo. Combining Segmentation and Classification Techniques for Fuzzy Knowledge-based Semantic Image Annotation. In *Proceedings of 3rd International Conference on Semantics And digital Media Technologies (SAMT 2008)*, December 2008, Koblenz, Germany.
3. E. Galmar, Th. Athanasiadis, B. Huet, Y. Avrithis. Spatiotemporal Semantic Video Segmentation. In *Proceedings of 10th International Workshop on Multimedia Signal Processing (MMSP 2008)*, October 2008, Cairns, Australia.
4. N. Simou, Th. Athanasiadis, S. Kollias, G. Stamou, A. Stafylopatis. Semantic Adaptation of Neural Network Classifiers in Image Segmentation. In *Proceedings of 18th International Conference on Artificial Neural Networks (ICANN 2008)*, Springer, pp. 907-916, September 2008, Prague, Czech Republic.
5. N. Simou, Th. Athanasiadis, S. Kollias. An Architecture for Multimedia Analysis and Retrieval Based on Fuzzy Description Logics. In *Proceedings of 2nd K-Space PhD Students Workshop, CEUR-WS*, July 2008, Paris, France.
6. N. Simou, Th. Athanasiadis, V. Tzouvaras, S. Kollias. Multimedia Reasoning with f-SHIN. In *Proceedings of 2nd International Workshop on Semantic Media Adaptation and Personalization (SMAP 2007)*, IEEE, pp. 44-49, December 2007, London, UK.
7. Th. Athanasiadis, S. Kollias. A Graph Based, Semantic Region Growing Approach in Image Segmentation. In *Proceedings of 1st K-Space PhD Students Workshop*, September 2007, Berlin, Germany.
8. Th. Athanasiadis, Ph. Mylonas, Y. Avrithis. A Context-based Region Labeling Approach for Semantic Image Segmentation. In *Proceedings of 1st International Conference on Semantics And digital Media Technology (SAMT 2006)*, Springer, pp. 212-225, December 2006, Athens, Greece.

9. P. Berka, Th. Athanasiadis, Y. Avrithis. Rule-based Reasoning for Semantic Image Segmentation and Interpretation. In Proceedings of 1st International Conference on Semantics And digital Media Technology (SAMT 2006), CEUR-WS, pp. 39-40, December 2006, Athens, Greece.
10. Ph. Mylonas, Th. Athanasiadis, Y. Avrithis. Image Analysis Using Domain Knowledge and Visual Context. In Proceedings of 13th International Conference on Systems, Signals and Image Processing (IWSSIP 2006), September 2006, Budapest, Hungary.
11. Th. Athanasiadis, Y. Avrithis, S. Kollias. A Semantic Region Growing Approach in Image Segmentation and Annotation. In Proceedings of 1st International Workshop on Semantic Web Annotations for Multimedia (SWAMM 2006), in conjunction with WWW, May 2006, Edinburgh, UK.
12. Ph. Mylonas, Th. Athanasiadis, Y. Avrithis. Improving Image Analysis using a Contextual Approach. In Proceedings of 7th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2006), April 2006, Seoul, Korea.
13. Th. Athanasiadis, V. Tzouvaras, K. Petridis, F. Precioso, Y. Avrithis, Y. Kompatsiaris. Using a Multimedia Ontology Infrastructure for Semantic Annotation of Multimedia Content. In Proceedings of 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot 2005), in conjunction with ISWC, CEUR-WS, pp. 59-68, November 2005, Galway, Ireland.
14. K. Petridis, F. Precioso, Th. Athanasiadis, Y. Avrithis, Y. Kompatsiaris. Combined Domain Specific and Multimedia Ontologies for Image Understanding. In Proceedings of 28th German Conference on Artificial Intelligence (KI 2005), September 2005, Koblenz, Germany.
15. N. Voisine, S. Dasiopoulou, V. Mezaris, E. Spyrou, Th. Athanasiadis, I. Kompatsiaris, Y. Avrithis, M. G. Strintzis. Knowledge-Assisted Video Analysis Using A Genetic Algorithm. In Proceedings of 6th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2005), April 2005, Montreux, Switzerland.
16. M. Wallace, Th. Athanasiadis, Y. Avrithis, G. Stamou, S. Kollias. A mediator system for hetero-lingual audiovisual content. In Proceedings of International Conference on Multi-platform e-Publishing, November 2004, Athens, Greece.
17. Th. Athanasiadis, Y. Avrithis. Adding Semantics to Audiovisual Content: The FAETHON Project. In Proceedings of 3rd International Conference for Image and Video Retrieval (CIVR 2004), Springer, pp. 665-673, July 2004, Dublin, Ireland.
18. M. Wallace, Th. Athanasiadis, Y. Avrithis. Knowledge Assisted Analysis & Categorization for Semantic Video Retrieval. In Proceedings of 3rd International Conference for Image and Video Retrieval (CIVR 2004), Springer, pp. 555-563, July 2004, Dublin, Ireland.

B.4 Τεχνικές Αναφορές

1. Th. Athanasiadis. Technical Report on Image Segmentation Evaluation. Image, Video and Multimedia Laboratory, NTUA, June 2008.
2. Th. Athanasiadis, E. Galmar. Technical Report on Spatiotemporal Semantic Segmentation of Video Shots. Multimedia Department, Eurecom, Sophia-Antipolis, France, February 2007.

□

Βιβλιογραφία

- [1] T. Adamek, N. O'Connor, and N. Murphy. Region-based segmentation of images using syntactic visual features. In *Workshop on Image Analysis for Multimedia Interactive Services, (WIAMIS), Montreux, Switzerland, 2005*.
- [2] R. Adams and L. Bischof. Seeded region growing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(6):641–647, June 1994.
- [3] J. F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.
- [4] R. Arnheim. *Art and Visual Perception*. University of California Press, new version; expanded and revised edition of the 1954 original edition, September 1974.
- [5] J. Assfalg, M. Bertini, C. Colombo, and A. Del Bimbo. Semantic annotation of sports videos. *IEEE Multimedia*, 9(2):52–60, 2002.
- [6] Th. Athanasiadis, Y. Avrithis, and S. Kollias. A semantic region growing approach in image segmentation and annotation. In *in Proceedings of 1st International Workshop on Semantic Web Annotations for Multimedia, in conjunction with WWW (SWAMM 2006)*, Edinburgh, UK, May 2006.
- [7] Th. Athanasiadis and S. Kollias. A graph based, semantic region growing approach in image segmentation. In *in Proceedings of 1st K-Space PhD Students Workshop*, Berlin, Germany, September 2007.
- [8] Th. Athanasiadis, Ph. Mylonas, and Y. Avrithis. A context-based region labeling approach for semantic image segmentation. In *in Proceedings of 1st International Conference on Semantics And digital Media Technology (SAMT 2006)*, pages 212–225, Athens, Greece, December 2006. Springer.
- [9] Th. Athanasiadis, Ph. Mylonas, Y. Avrithis, and S. Kollias. Semantic image segmentation and object labeling. *IEEE Trans. on Circuits and Systems for Video Technology*, 17(3):298–312, 2007.
- [10] Th. Athanasiadis, N. Simou, G. Papadopoulos, R. Benmokhtar, K. Chandramouli, V. Tzouvaras, V. Mezaris, M. Phinikettos, Y. Avrithis, Y. Kompatsiaris, B. Huet, and E. Izquierdo. Integrating image segmentation and classification for fuzzy knowledge-based multimedia indexing. In *in Proceedings of 15th International MultiMedia Modeling Conference (MMM 2009)*, pages 263–274, Sophia Antipolis, France, January 2009. Springer.

- [11] F. Baader, D. McGuinness, D. Nardi, and P.F. Patel-Schneider. *The description logic handbook: theory, implementation, and applications*. Cambridge University Press, 2003.
- [12] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein eds. OWL web ontology language reference. Technical report, World Wide Web Consortium, Feb 2004.
- [13] D. Beckett and B. McBride. RDF/XML syntax specification (revised). Technical report, W3C, 2004.
- [14] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 284(5):34–43, 2001.
- [15] J. C. Bezdek, J. Keller, R. Krisnapuram, and M. R. Pal. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Kluwer Academic Publishers, Norwell, MA, USA, 1999.
- [16] J.C. Bezdek. *Pattern recognition with fuzzy objective function algorithms*. Kluwer Academic Publishers Norwell, MA, USA, 1981.
- [17] C.M. Bishop. *Pattern recognition and machine learning*. Springer, 2007.
- [18] M. Bober. Mpeg-7 visual shape descriptors. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(6), June 2001.
- [19] J.S. Boreczky and L.A. Rowe. Comparison of video shot boundary detection techniques. *Journal of Electronic Imaging*, 5(2):122–128, 1996.
- [20] E. Borenstein, E. Sharon, and S. Ullman. Combining top-down and bottom-up segmentation. In E. Dougherty, editor, *Proceedings of IEEE workshop on Perceptual Organization in Computer Vision, CVPR 2004*, 2004.
- [21] E. Borenstein and S. Ullman. Class-Specific, Top-Down Segmentation. In *In Proc. of 7th European Conference on Computer Vision, ECCV*. Springer, 2002.
- [22] M.R. Boutell, J. Luo, X. Shen, and C.M. Brown. Learning multi-label scene classification. *Pattern Recognition*, 37(9):1757–1771, 2004.
- [23] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1222–1239, 2001.
- [24] D. Brickey and RV Guha. Resource Description Framework (RDF) Schema specification 1.0. Technical report, W3C, March 2000.
- [25] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1026–1038, 2002.

- [26] HD Cheng, XH Jiang, Y. Sun, and J. Wang. Color image segmentation: advances and prospects. *Pattern Recognition*, 34(12):2259–2281, 2001.
- [27] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):603–619, 2002.
- [28] Dublin Core Community. Dublin core element set. *Orion Business*, Version 1.1, 1999. <http://www.dublincore.org/documents/dces/>.
- [29] J. Cousty, G. Bertrand, L. Najman, and M. Couprie. Watershed cuts: minimum spanning forests and the drop of water principle. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(8):1362–1374, 2009.
- [30] D. Crever and R. Lepage. Knowledge-Based Image Understanding Systems: A Survey. *Computer Vision and Image Understanding*, 67(2):161–185, 1997.
- [31] S. Dasiopoulou, V. Mezaris, I. Kompatsiaris, V.-K. Papastathis, and M. G. Strintzis. Knowledge-assisted semantic video object detection. *IEEE Trans. on Circuits and Systems for Video Technology*, 15(10):1210–1224, 2005.
- [32] D. DeMenthon and D. Doermann. Video retrieval using spatio-temporal descriptors. In *Proceedings of the eleventh ACM international conference on Multimedia*, pages 508–517. ACM, 2003.
- [33] E. Di Sciascio, F. M. Donini, and M. Mongiello. Structured knowledge representation for image retrieval. *Journal of Artificial Intelligence Research (JAIR)*, 16:209–257, 2002.
- [34] R. Diestel. *Graph Theory*. Springer, Heidelberg, 3rd edition, 2005.
- [35] H. Digabel and C. Lantuejoul. Iterative algorithms. In *Proc. 2nd European Symp. Quantitative Analysis of Microstructures in Material Science, Biology and Medicine, Caen, France, Oct*, pages 85–99, 1977.
- [36] M.A. Eshera and K.-S. Fu. An image understanding system using attributed symbolic representation and inexact graph-matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(5):604–618, 1986.
- [37] J. Fan, D.K.Y. Yau, A.K. Elmagarmid, and W.G. Aref. Automatic image segmentation by integrating color-edge extraction and seeded region growing. *Image Processing, IEEE Transactions on*, 10(10):1454–1466, October 2001.
- [38] J.A. Feldman and Y. Yakimovsky. Decision theory and artificial intelligence: A semantics-based region analyzer. *Artificial Intelligence*, 5(4):349–371, 1974.
- [39] P.F. Felzenszwalb and D.P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.
- [40] E. Galmar, Th. Athanasiadis, B. Huet, and Y. Avrithis. Spatiotemporal semantic video segmentation. In *in Proceedings of 10th International Workshop on Multimedia Signal Processing (MMSP 2008)*, Cairns, Australia, October 2008. IEEE.

- [41] E. Galmar and B. Huet. Graph-based spatio-temporal region extraction. *Lecture Notes in Computer Science*, 4141:236, 2006.
- [42] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. Sweetening ontologies with dolce. *Lecture notes in computer science*, pages 166–181, 2002.
- [43] F. Ge, S. Wang, and T. Liu. Image-segmentation evaluation from the perspective of salient object extraction. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2006.
- [44] F. Ge, S. Wang, and T. Liu. New benchmark for image segmentation evaluation. *Journal of Electronic Imaging*, 16(3):033011–1, 2007.
- [45] L. Goldmann, T. Adamek, P. Vajda, M. Karaman, R. Mörzinger, E. Galmar, T. Sikora, N. O’Connor, T. Ha-Minh, T. Ebrahimi, P. Schallauer, and B. Huet. Towards fully automatic image segmentation evaluation. In *submitted*, 2008.
- [46] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, 3rd edition, August 2007.
- [47] H. Greenspan, J. Goldberger, and A. Mayer. Probabilistic space-time video modeling via piecewise GMM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):384–396, 2004.
- [48] T.R. Gruber et al. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220, 1993.
- [49] P. Hayes. Rdf semantics. Technical report, W3C, February 2004.
- [50] A. Hoover, G. Jean-Baptiste, X. Jiang, P.J. Flynn, H. Bunke, D.B. Goldgof, K. Bowyer, D.W. Eggert, A. Fitzgibbon, and R.B. Fisher. An experimental comparison of range image segmentation algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(7):673–689, Jul 1996.
- [51] I. Horrocks, U. Sattler, and S. Tobies. Reasoning with individuals for the description logic shiq. In David MacAllester, editor, *CADE-2000*, number 1831 in LNAI, pages 482–496. Springer-Verlag, 2000.
- [52] Q. Huang and B. Dom. Quantitative methods of evaluating image segmentation. In *IEEE International Conference on Image Processing*, 1995.
- [53] J. Hunter, J. Drennan, and S.Little. Realizing the hydrogen economy through semantic web technologies. *IEEE Intelligent Systems*, 19(1):40–47, 2004.
- [54] ISO/IEC. Overview of the MPEG-7 Standard (version 6.0). ISO/IEC JTC1/SC29/WG11/N4980, Pattaya, December 2001.
- [55] ISO/IEC. MPEG-21 Overview v.5. ISO/IEC JTC1/SC29/WG11/N5231, Shanghai, October 2002.
- [56] A. Jaimes and J. R. Smith. Semi-automatic, data-driven construction of multimedia ontologies. *Proc. IEEE Intl. Conf. on Multimedia and Expo (ICME)*, March 2003.

- [57] AK Jain, MN Murty, and PJ Flynn. Data clustering: a review. *ACM computing surveys*, 31(3), 1999.
- [58] G. J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, New Jersey, NJ, USA, 1995.
- [59] I. Kokkinos and P. Maragos. Synergy Between Object Recognition and Image Segmentation using the Expectation Maximization Algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(8):1486–1501, 2009.
- [60] S.H. Kwok, A.G. Constantinides, and W.C. Siu. An efficient recursive shortest spanning tree algorithm using linking properties. *IEEE Transactions on circuits and systems for video technology*, 14(6):852–863, 2004.
- [61] S. Lee and MM Crawford. Unsupervised classification using spatial region growing segmentation and fuzzy training. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 1, 2001.
- [62] B. Leibe and B. Schiele. Interleaved object categorization and segmentation. In *In Proc. of British Machine Vision Conference*, pages 759–768, 2003.
- [63] Y. Li, J. Sun, and H.Y. Shum. Video object cut and paste. *ACM Transactions on Graphics (TOG)*, 24(3):600, 2005.
- [64] X. Liu, A. Srivastava, and K. Gallivan. Optimal linear representations of images for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):662–666, 2004.
- [65] J. Lladós, E. Martí, and J.J. Villanueva. Symbol recognition by error-tolerant subgraph matching between region adjacency graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1137–1143, 2001.
- [66] J. Luo and A. Savakis. Indoor vs outdoor classification of consumer photographs using low-level and semantic features. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 2, 2001.
- [67] N. Maillot and M. Thonnat. A weakly supervised approach for semantic image indexing and retrieval. *Lecture notes in computer science*, 3568:629, 2005.
- [68] N. Maillot, M. Thonnat, and A. Boucher. Towards ontology-based cognitive vision. *Machine Vision and Applications*, 16(1):33–40, 2004.
- [69] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. In *Special Issue on MPEG-7, IEEE Trans. on Circuits and Systems for Video Technology*, 11/6, 703-715, 2001.
- [70] P. Maragos. *Notes on Computer Vision*. NTUA, 2005.
- [71] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman, January 1982.
- [72] D. Martin. *An empirical approach to grouping and segmentation*. PhD thesis, University of California, Berkeley, 2002.

- [73] 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 *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 2001.
- [74] K. McGuinness, G. Keenan, T. Adamek, and N. O'Connor. Image segmentation evaluation using an integrated framework. In *VIE 2007 - Proceedings of the IET 4th International Conference on Visual Information Engineering 2007*, 2007.
- [75] C. Meghini, F. Sebastiani, and U. Straccia. Reasoning about the form and content for multimedia objects (extended abstract). In *Proceedings of AAAI 1997 Spring Symposium on Intelligent integration and Use of Text, Image, Video and Audio*, pages 89–94, Stanford University, California, 1997.
- [76] C. Meghini, F. Sebastiani, and U. Straccia. A model of multimedia information retrieval. *Journal of the ACM*, 48(5):909–970, 2001.
- [77] F. Meyer and S. Beucher. Morphological segmentation. *Journal of Visual Communication and Image Representation*, 1(1):21–46, September 1990.
- [78] C. Millet, I. Bloch, P. Hede, and PA Moellic. Using relative spatial relationships to improve individual region recognition. In *Proceedings of 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies*, pages 119–126, 2002.
- [79] J.W. Modestino and J. Zhang. A markov random field model-based approach to image interpretation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(6):606–615, 1992.
- [80] R. Moeller, B. Neumann, and M. Wessel. Towards computer vision with description logics-some recent progress. In *Workshop on Integration of Speech and Image Understanding*, pages 101–115, 1999.
- [81] R. Moller and B. Neumann. Ontology-Based Reasoning Techniques for Multimedia Interpretation and Retrieval. *Semantic multimedia and ontologies: theory and applications*, page 55, 2008.
- [82] O. J. Morris, M. J. Lee, and A. G. Constantinides. Graph theory for image analysis: An approach based on the shortest spanning tree. In *Inst. Elect. Eng., vol. 133, pp. 146-152*, 1986.
- [83] B. Motik. On the properties of metamodeling in owl. *Journal of Logic and Computation*, 17(4):617, 2007.
- [84] PJ Mulroy. Video content extraction: Review of current automatic segmentation algorithms. In *Proc. Workshop Image Analysis and Multimedia Interactive Systems (WIAMIS)*, 1997.
- [85] P. Mylonas. *Adaptive Multimedia Content Access based on Context and User Profiles*. PhD thesis, National Technical University of Athens, Athens, June 2006.

- [86] Ph. Mylonas, Th. Athanasiadis, M. Wallace, Y. Avrithis, and S. Kollias. Semantic representation of multimedia content - knowledge representation and semantic indexing. *Springer Multimedia Tools and Applications*, 39(3):293–327, September 2008.
- [87] F. Nack, J. van Ossenbruggen, and L. Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part II). *IEEE Multimedia*, 12(1), 2005. based on <http://ftp.cwi.nl/CWIREports/INS//INS-E0309.pdf>.
- [88] HR Naphade and TS Huang. A probabilistic framework for semantic video indexing, filtering, and retrieval. *IEEE Transactions on Multimedia*, 3(1):141–151, 2001.
- [89] M.R. Naphade, I. Kozintsev, and T.S. Huang. A Factor graph framework for semantic video indexing. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(1):40–52, 2002.
- [90] B. Neumann and R. Moeller. On scene interpretation with description logics. In *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, number 3948 in Springer, pages 247–278, 2004.
- [91] University of Karlsruhe. The karlsruhe ontology and semantic web tool suite.
- [92] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, 1979.
- [93] N.R. Pal and S.K. Pal. A review on image segmentation techniques. *Pattern recognition*, 26(9):1277–1294, 1993.
- [94] N. Papamarkos, C. Strouthopoulos, and I. Andreadis. Multithresholding of color and gray-level images through a neural network technique. *Image and Vision Computing*, 18(3):213–222, 2000.
- [95] P. F. Patel-Schneider, P. Hayes, and I. Horrocks. Owl web ontology language semantics and abstract syntax. Technical report, W3C, February 2004.
- [96] S.E. Peraldi, A. Kaya, S. Melzer, R. Moller, and M. Wessel. Multimedia Interpretation as Abduction.
- [97] WA Perkins. Area segmentation of images using edge points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:8–15, 1980.
- [98] M.A. Peterson. Object recognition processes can and do operate before figure-ground organization. *Current Directions in Psychological Science*, pages 105–111, 1994.
- [99] J.M. Prager. Extracting and labeling boundary segments in natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(1):16–27, 1980.
- [100] The Mandarax Project. The mandarax deduction rules engine.

- [101] K. Rapantzikos, Y. Avrithis, and S. Kollias. Dense saliency-based spatiotemporal feature points for action recognition. In *in Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, Miami, USA, June 2009.
- [102] RuleML. The rule markup initiative.
- [103] S.J. Russell, P. Norvig, J.F. Canny, J. Malik, and D.D. Edwards. *Artificial intelligence: a modern approach*. Prentice hall Englewood Cliffs, NJ, 1995.
- [104] P. Salembier and L. Garrido. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE transactions on Image Processing*, 9(4):561–576, 2000.
- [105] P. Salembier and F. Marques. Region-based representations of image and video: Segmentation tools for multimedia services. *Circuits and Systems for Video Technology, IEEE Transactions on*, 9(8):1147–1169, 1999.
- [106] P. Salembier and J.R. Smith. Mpeg-7 multimedia description schemes. *IEEE transactions on circuits and systems for video technology*, 11(6):748–759, 2001.
- [107] Hermes-T. Herzog O. Schober, J.P. Content-based image retrieval by ontology-based object recognition. In *In Proc. of 27th German Conference on Artificial Intelligence*, September 2004.
- [108] E. Di Sciascio and F. Donini. Description logics for image recognition: a preliminary proposal. In *Proceedings of the International Workshop on Description Logics (DL 99)*, 1999.
- [109] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.
- [110] J. Shotton, M. Johnson, R. Cipolla, T.C.R.D. Center, and J. Kawasaki. Semantic texton forests for image categorization and segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
- [111] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *European Conference on Computer Vision*. Springer, 2006.
- [112] T. Sikora. The MPEG-7 Visual standard for content description - an overview. *IEEE Trans. on Circuits and Systems for Video Technology, special issue on MPEG-7*, 11(6):696–702, June 2001.
- [113] N. Simou, Th. Athanasiadis, and S. Kollias. An architecture for multimedia analysis and retrieval based on fuzzy description logics. In *in Proceedings of 2nd K-Space PhD Students Workshop*, Paris, France, July 2008. CEUR-WS.
- [114] N. Simou, Th. Athanasiadis, S. Kollias, and G. Stamou. Semantic adaptation of neural network classifiers in image segmentation. *Neural Network World*, 19(5):561–579.

- [115] N. Simou, Th. Athanasiadis, S. Kollias, G. Stamou, and A. Stafylopatis. Semantic adaptation of neural network classifiers in image segmentation. In *in Proceedings of 18th International Conference on Artificial Neural Networks (ICANN 2008)*, pages 907–916, Prague, Czech Republic, September 2008. Springer.
- [116] N. Simou, Th. Athanasiadis, G. Stoilos, and S. Kollias. Image indexing and retrieval using expressive fuzzy description logics. *Signal, Image and Video Processing*, 2(4):321–335, December 2008.
- [117] N. Simou, Th. Athanasiadis, V. Tzouvaras, and S. Kollias. Multimedia reasoning with f-shin. In *in Proceedings of 2nd International Workshop on Semantic Media Adaptation and Personalization (SMAP 2007)*, pages 44–49, London, UK, December 2007. IEEE.
- [118] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1349–1380, 2000.
- [119] J. R. Smith and S.-F. Chang. Visualseek: A fully automated content-based image query system. In *Proceedings of the 4th ACM international conference on Multimedia, Boston, USA, 1997*.
- [120] A. Sofou. *Image Segmentation Using Morphological Methods and Active Contours with Application to Soilsection Analysis*. PhD thesis, National Technical University of Athens, Athens, June 2006.
- [121] M Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis and Machine Vision*. CL-Engineering, 3rd edition, March 2007.
- [122] G. Stamou and S. Kollias. *Multimedia Content and the Semantic Web: Methods, Standards and Tools*. John Wiley & Sons Ltd, 2005.
- [123] G. Stamou, J. van Ossenbruggen, J. Pan, and G. Schreiber. Multimedia annotations on the semantic web. *IEEE Multimedia*, 13(1):86–90, January – March 2006.
- [124] G. Stoilos. *Uncertainty and the Semantic Web: Expressive Fuzzy Description Logics*. PhD thesis, National Technical University of Athens, Athens, June 2006.
- [125] G. Stoilos, G. Stamou, and J. Z. Pan. Handling imprecise knowledge with fuzzy description logics. In *Proceedings of the International Workshop on Description Logics (DL 2006), Lake District, UK, 2006*.
- [126] G. Stoilos, G. Stamou, V. Tzouvaras, J.Z. Pan, and I. Horrocks. The fuzzy description logic f-*SHLN*. In *Proc. of the International Workshop on Uncertainty Reasoning for the Semantic Web, 2005*.
- [127] G. Stoilos, G. Stamou, V. Tzouvaras, J.Z. Pan, and I. Horrocks. Fuzzy OWL: Uncertainty and the semantic web. In *Proc. of the International Workshop on OWL: Experiences and Directions, 2005*.

- [128] U. Straccia. Reasoning within fuzzy description logics. *Journal of Artificial Intelligence Research*, 14:137–166, 2001.
- [129] T.M. Strat and M. Fischler. Context-based vision: recognizing objects using information from both 2D and 3D imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):1050–1065, 1991.
- [130] G. Papadopoulos V. Mezaris Y. Avrithis I. Kompatsiaris M. Strintzis Th. Athanasiadis, Ph. Mylonas. *Knowledge Driven Segmentation and Classification. In Multimedia Semantics: Metadata, Analysis and Interaction*, chapter eleven. Willey, in press.
- [131] A. Torralba. Contextual priming for object detection. *International Journal of Computer Vision*, 53(2):169–191, 2003.
- [132] R. Troncy. Integrating Structure and Semantics into Audio-visual Documents. In *Proc. of the International Workshop on Uncertainty Reasoning for the Semantic Web*, pages 566 – 581. Springer-Verlag Heidelberg, 2003.
- [133] W.-H. Tsai and K.-S. Fu. Error-correcting isomorphisms of attributed relational graphs for pattern analysis. *Systems, Man and Cybernetics, IEEE Transactions on*, 9(12):757–768, December 1979.
- [134] Z. Tu and S.-C. Zhu. Image segmentation by data-driven markov chain monte carlo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):657–673, May 2002.
- [135] R. Unnikrishnan, C. Pantofaru, and M. Hebert. Toward objective evaluation of image segmentation algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(6):929–944, June 2007.
- [136] J. van Ossenbruggen, F. Nack, and L. Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part I). *IEEE Multimedia*, 11(4):38–48, October – December 2004. based on <http://ftp.cwi.nl/CWIreports/INS//INS-E0308.pdf>.
- [137] J. van Ossenbruggen, R. Troncy, G. Stamou, and J. Z. Pan. Image annotation on the semantic web. Technical report, W3C, March 2006.
- [138] V.N. Vapnik. *The nature of statistical learning theory*. Springer Verlag, 2000.
- [139] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE transactions on pattern analysis and machine intelligence*, 13(6):583–598, 1991.
- [140] I. Vlahavas, P. Kefalas, N. Bassiliades, F. Kokkoras, and I. Sakellariou. *Artificial Intelligence*. V.Giurdas Publications, Greece, 3rd edition, 2006.
- [141] W3C. Resource description framework (rdf).
- [142] D. B. West. *Introduction to Graph Theory*. Prentice Hall, 2nd edition, 2001.

- [143] P. Wu, B. S. Manjunath, S. Newsam, and H. D. Shin. A texture descriptor for browsing and similarity retrieval. In *Signal Processing: Image Communication, Volume 16*, pp. 33-43, 2000.
- [144] R. R. Yager and L. A. Zadeh, editors. *Fuzzy Sets, Neural Networks and Soft Computing*. John Wiley & Sons, Inc., New York, NY, USA, 1994.
- [145] S.X. Yu, R. Gross, and J. Shi. Concurrent Object Recognition and Segmentation by Graph Partitioning. In *Neural Information Processing Systems*, volume 1, 2002.
- [146] YJ Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.

□

Βιογραφικό Σημείωμα

Ο Θάνος Αθανασιάδης γεννήθηκε στην Καβάλα στις 5 Φεβρουαρίου 1980 και μεγάλωσε στην Αθήνα. Αποφοίτησε το 2003 από τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσοβίου Πολυτεχνείου και συνέχισε την εκπαίδευσή του ως υποψήφιος διδάκτορας στο εργαστήριο Ψηφιακής Επεξεργασίας Εικόνων, Βίντεο και Πολυμέσων υπό την επίβλεψη του καθηγητή Στέφανου Κόλλια. Στο πλαίσιο της διδακτορικής του διατριβής έχει δημοσιεύσει άρθρα σε επιστημονικά περιοδικά και συνέδρια, καθώς επίσης έχει συμμετάσχει σε ερευνητικά και αναπτυξιακά προγράμματα.

Σπουδές

Ph.D. “Τεχνολογίες Γνώσης και Πρόσβαση σε Οπτικοακουστικές Πληροφορίες”
Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνων, Βίντεο και Πολυμέσων, ΕΜΠ, 2009

- Υποτροφία ΠΕΝΕΔ, “Σημασιολογική Αναζήτηση Πολυμεσικού Υλικού με Χρήση Τεχνολογιών Γνώσης”, 2005-2009
- Θωμαΐδειο Βραβείο για την πρόοδο των επιστημών και των τεχνών, 2005, 2006, 2007

Δίπλωμα Ηλεκτρολόγου Μηχανικού και Μηχανικού Υπολογιστών, ΕΜΠ, 2003

- Κατεύθυνση: Ηλεκτρονικής και Συστημάτων
- Διπλωματική Εργασία: Εικονική Αναπαράσταση της Ελληνικής Νοηματικής Γλώσσας στο Διαδίκτυο

Ερευνητικό Προφίλ

Ερευνητικά Ενδιαφέροντα:

- Κατάτμηση και σημασιολογικός χαρακτηρισμός εικόνων και βίντεο
- Γραφοθεωρητικές μέθοδοι και ασαφής συνολοθεωρία με εφαρμογές στην ανάλυση εικόνων και βίντεο
- Αναπαράσταση και χρήση γνώσης για ανάλυση και περιγραφή πολυμεσικού περιεχομένου

Συμμετοχή σε 3 ευρωπαϊκά ερευνητικά προγράμματα (K-Space, aceMedia, FAETHON)

Συγγραφέας 6 άρθρων σε επιστημονικά περιοδικά, ενός κεφαλαίου βιβλίου και 18 άρθρων σε συνέδρια με κριτές

Εξωτερικός κριτής σε επιστημονικά περιοδικά και συνέδρια

Συνδιοργανωτής και μέλος επιτροπής προγράμματος συνεδρίων και ημερίδων

Διδακτική Εμπειρία

Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων:

- Ψηφιακή Επεξεργασία Εικόνας και Τεχνολογίες Βίντεο, 2003-2008
- Γραφική με Υπολογιστές, 2003-2005

Primus IT Seminars: Εισηγητής σεμιναρίων MS Excel, MS Access, 2004

Moscow Aviation Institute, Department of Computational Mathematics: Υλοποίηση εκπαιδευτικού λογισμικού για επίλυση διαφορικών εξισώσεων, 2002

Προσωπικά Στοιχεία

Ξένες γλώσσες: Αγγλικά, Γερμανικά, Ισπανικά

Μέλος TEE (2004), IEEE (2004)

Γενικά ενδιαφέροντα:

- Αθλητικές δραστηριότητες: μπάσκετ, κολύμβηση
- Ταξίδια, γνωριμία με διαφορετικές κουλτούρες και πολιτισμούς
- Εθελοντικές δράσεις για το περιβάλλον και τον άνθρωπο

□